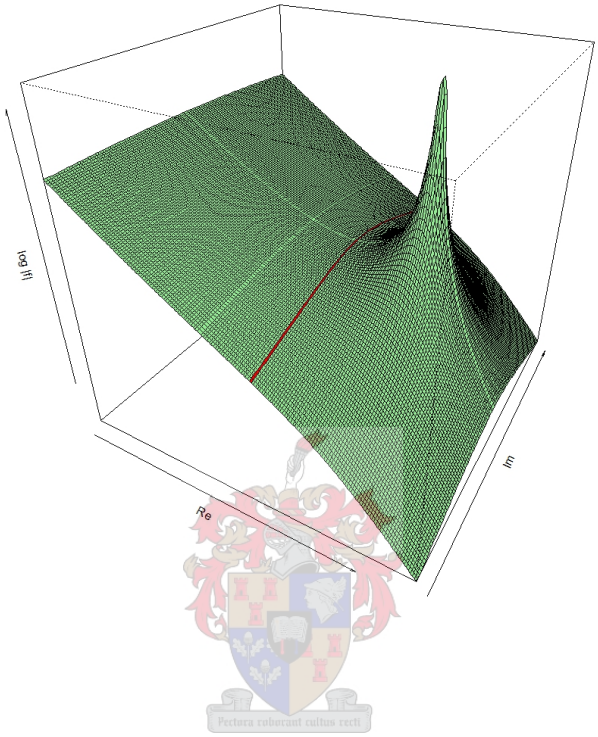
The Saddle-point Method and its Application to the Hill Estimator

Sven Buitendag



Thesis presented in partial fulfilment
of the requirements for the degree of
MComm (Mathematical Statistics)
at Stellenbosch University

Supervisor: Professor Tertius de Wet

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Department of Statistics and Actuarial Science
Stellenbosch University
South Africa
December, 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: Sven Buitendag

Date: December 2016

Abstract

The saddle-point approximation is a highly accurate approximation of the distribution of a random variable. It was originally derived as an approximation in situations where a parameter takes on large values. However, due to its high accuracy and good behaviour in a variety of applications not involving such a parameter, it has been generalised and applied to the distribution of any random variable with a well-behaved cumulant generating function.

In this thesis the theory underlying the saddle-point approximation will be discussed and illustrated with an application to approximate the distribution of the Hill estimator in extreme value theory.

Opsomming

Die saalpunt-benadering is 'n hoogs noukeurige benadering van die verdeling van 'n stochastiese veranderlike. Dit is oorspronklik afgelei as 'n benadering in gevalle waar 'n parameter groot waardes aanneem. Nietemin, na aanleiding van sy hoogs akkurate aard en goeie gedrag in 'n verskeidenheid van toepassings wat nie betrekking het op so 'n parameter nie, is dit veralgemeen en toegepas op die verdeling van enige stochastiese veranderlike met 'n kumulant-voortbringende funksie wat goeie gedrag toon.

In hierdie tesis sal die teorie onderliggend aan die saalpunt-benadering bespreek en gellustreer word met 'n toepassing om die verdeling van die Hill-beramer te benader.

Acknowledgements

I would like to thank my supervisor, Professor Tertius de Wet, for the time and effort he invested in guiding me in this research task. I would also like to thank Stellenbosch University, namely the Library and the Department of Statistics and Actuarial Science, for providing access to international statistical and mathematical journals.

Contents

Chapter 1

Introduction

1.1 The Purpose of this Research

1.1.1 An Overview of the Thesis

The focus of this thesis is around the saddle-point method, the saddle-point approximation, and its application to the Hill estimator. The saddle-point method is an approach to deriving an asymptotic estimate of a complex integral. The saddle-point method yields the saddle-point density and distribution functions when applied to the Fourier inversion formula of a probability density and distribution function respectively. The saddle-point approximation refers to both the saddle-point density and distribution functions, and is perhaps best summarised by the opening line of the book by Butler (2007): "Among the various tools that have been developed for use in statistics and probability over the years, perhaps the least understood and most remarkable tool is the saddlepoint approximation."

The saddle-point approximation historically refers to the uniform asymptotic expansion of the distribution of certain random variables which depend on a large parameter, such as the sample size. It became apparent, however, that the saddle-point approximation serves as an accurate approximation of the distribution of random variables that do not depend on large parameters, including the sample size (see Hougaard (1988)). "It is remarkable because it usually provides probability approximations whose accuracy is much greater than the current supporting theory would suggest" (Butler, 2007).

This thesis has three central topics: the saddle-point method, the accuracy of the saddle-point approximation in a non-asymptotic setting and the good behaviour of the saddle-point density and distribution functions of the Hill estimator.

The saddle-point method is the theory which underlies the the saddle-point approximation. The saddle-point approximation of Daniels (1954) and Lugannani & Rice (1980) are direct applications of the saddle-point methods of de Bruijn (1958) and Bleistein (1966) respectively. The link between the saddle-point approximation of Daniels (1954) and the saddle-point method of de Bruijn (1958) only became apparent after the book of de Bruijn (1958) was published. An overview of the saddle-point method is given in the second chapter of this thesis.

The saddle-point approximation is very accurate in many non-asymptotic settings although there is little or no theory supporting this. This is because the theory which underlies the saddle-point approximation, namely the saddle-point method, is asymptotic. It therefore is necessary to adapt the asymptotic theory of the saddle-point method for it to yield a non-

asymptotic saddle-point approximation. Although no rigorous derivation of such an adaptation is provided in this thesis, the differences in the asymptotic saddle-point method and the non-asymptotic saddle-point approximation are discussed, and possible adaptations to the asymptotic saddle-point method are proposed. The third chapter in this thesis discusses the saddle-point approximation.

The saddle-point density and distribution functions of the Hill estimator are very well behaved, even in cases when the parameter values are unknown and have to be estimated. The Hill estimator, its saddle-point approximation and estimation thereof are all discussed in the fourth chapter of this thesis.

1.1.2 The Aim of the Thesis

This thesis aims to contribute to the field of statistics in the following four ways:

Firstly, the thesis provides a coherent discussion on how the saddle-point method is applied to the inverse Fourier transform to yield the saddle-point approximation.

Secondly, the thesis addresses the need for theory to be developed regarding the saddle-point approximation in a non-asymptotic setting. The saddle-point approximation is an asymptotic result, but due to its accuracy in a variety of non-asymptotic applications, its use has become popular in the non-asymptotic setting. In order to establish itself in the non-asymptotic setting, however, it is necessary to develop or adapt the theory underlying the method for this setting. Such adaptions are proposed in the third chapter.

Thirdly, the thesis proposes saddle-point approximations for the density and distribution functions of the Hill estimator. These saddle-point approximations prove to be highly accurate for a variety of heavy-tailed distributions.

Finally, the thesis illustrates the good behaviour of the saddle-point approximation for unstable parameter values. The saddle-point approximation of the Hill estimator does not only yield accurate approximations when the actual parameter values are known, but also yields well-behaved approximations when the parameter values are estimated using a small sample and are therefore highly unstable. This is illustrated by a simulation study, which is discussed in the fourth chapter.

1.2 Outline of the Thesis

The first chapter is introductory and provides a brief discussion of the meaning, history and applications of both the saddle-point method and the saddle-point approximation. This chapter is structured so that a section is dedicated to the saddle-point method and another section is dedicated to the saddle-point approximation.

The second chapter provides an overview of the saddle-point method. This overview includes an introductory section on the complex analysis that concerns the saddle-point method and a discussion of the derivation of the method and its extensions. This chapter lays the groundwork for the saddle-point method which is applied in the subsequent chapter.

The third chapter discusses the saddle-point approximation, with the focus on its general definition and applications as given by Butler (2007). In this chapter, the theory of the saddle-point method is applied (and partially adapted) to the Fourier inversion formula to yield the general saddle-point density and distribution functions of Butler (2007). The chapter is structured so that a section is dedicated to the saddle-point density function and another section is dedicated to the saddle-point distribution function. One of the primary aims of this chapter is to address the issue that the theory underlying the saddle-point approximation is asymptotic, while the saddle-point approximation itself can be applied generally, including in a non-asymptotic setting. Furthermore, the chapter includes a discussion of the saddle-point density and distribution functions in their original forms as given by Daniels (1954) and Lugannani & Rice (1980) respectively.

In the fourth chapter, saddle-point density and distribution functions of the Hill estimator are proposed. These saddle-point density and distribution functions rely on a second-order approximation regarding the distribution of the Hill estimator. This chapter provides an introductory section on extreme value theory and the Hill estimator, after which the second-order approximation regarding the Hill estimator is discussed and finally the saddle-point approximation is applied to yield the saddle-point density and distribution functions of the Hill estimator. The accuracy of the saddle-point density and distribution functions are illustrated for a variety of heavy-tailed distributions. Furthermore, the accuracy and behaviour of the saddle-point density and distribution functions are investigated when the parameters are unknown and therefore need to be estimated.

The final chapter serves as a conclusion. All the key findings of the thesis are summarised in this chapter. The conclusion is structured around the three central topics discussed in section 1.1 and also includes discussions of other key findings that are revealed throughout the thesis.

1.3 The Saddle-point Method

1.3.1 Introduction

The saddle-point method is an approach to deriving an asymptotic estimate of a complex integral. The method was formally introduced by Debeye in 1909 but was previously derived by Riemann in unpublished work in 1863 and by a Russian mathematician Nekrasov in his doctoral dissertation in 1885 (Petrova & Solov'ev, 1997:1). Since its introduction, the saddle-point method has been used extensively in mathematics and applied mathematics.

This section has two parts. Firstly, the history of the saddle-point method is briefly discussed in section 1.3.2. This discussion follows closely that of Petrova & Solov'ev (1997) Secondly, some of the applications of the saddle-point method are discussed in section 1.3.3. The theory underlying the saddle-point method is discussed later in chapter 2.

1.3.2 A Brief History of the Saddle-point Method in Mathematics

The saddle-point method, also called the method of steepest descent, was first formally introduced by Debeye in 1909, who developed the method in his study of the asymptotics of Bessel functions. Debeye admitted that he had gotten the idea from an unpublished paper by Riemann, who developed the method to estimate the asymptotic behaviour of the hypergeometric function in 1863. The method, however, can be traced further back to Cauchy, who developed a basic saddle-point method in 1827 in his study of the radius of convergence of the Lagrange series. The method of Cauchy did not hold generally since it " always took the contour to be a circle, thus depriving his method of the necessary generality" (Petrova & Solov'ev, 1997:367).

In Riemann's unpublished paper, which was written in 1863, he uses the saddle-point method to find an asymptotic expansion of an integral which is related to the hypergeometric function. This work of Riemann was found by Schwarz, who was subsequently published it after making "the necessary annotations to the fragments left behind by Riemann at his death" (Petrova & Solov'ev, 1997:368). Further evidence that Riemann had developed the saddle-point method is provided by his other unpublished papers in which he estimates the function $Z(t)$ (which is connected with the Riemann zeta-function). These papers were found and published by Siegel in 1932 (Petrova & Solov'ev, 1997:369).

The Russian mathematician Pavel Alexeevich Nekrasov developed the saddle-point method in his 1885 doctoral dissertation in which he studied the asymptotics of the radius of convergence of the Lagrange series. In his dissertation, Nekrasov "devotes an especially large amount of attention to the paper of Cauchy (of 1827)". (Petrova & Solov'ev, 1997:370). In the third chapter of his dissertation, which was published in 1885, Nekrasov developed the saddle-point approximation in "the most general case, when there are several saddle points and each has an arbitrary multiplicity." (Petrova & Solov'ev, 1997:374). The work of Nekrasov, however, largely went unnoticed by Western mathematicians for several years.

Debeye developed the method in 1909 in his study of the asymptotics of Bessel functions, which are functions that solve the Bessel differential equation (Petrova & Solov'ev, 1997). At the start of his paper, Debye notes that the idea of the saddle-point method came to him from reading the unpublished papers by Riemann which are discussed above.

1.3.3 Applications of the Saddle-point Method

This subsection discusses some of the existing applications of the saddle-point method in mathematics. The saddle-point method can be used to derive asymptotic estimates of solutions to differential equations. Such solutions include the hypergeometric functions, the Bessel functions and the Airy functions. The saddle-point method can also be used to estimate Cauchy coefficient estimates and the Riemann zeta function. A brief discussion of each of these applications is provided, the purpose being to emphasise the variety of ways in which the saddle-point method can be applied.

The Hypergeometric Functions

Riemann uses the saddle-point method to derive an asymptotic estimate of an integral which is related to the hypergeometric function (Petrova & Solov'ev, 1997:368-369). The hypergeometric functions are solutions to the hypergeometric differential equation

$$x\,(1-x)\,\frac{d^2y}{dx^2} + (c-(a+b+1)\,x)\,\frac{dy}{dx} - a\,b\,y = 0$$

where a, b, c are real and x can be complex. The so-called regular solution is given by

$$\begin{aligned} y &= 1 + \frac{ab}{c}\,x + \frac{a(a+1)b(b+1)}{2\,c(c+1)}\,x^2 + \frac{1}{3!}\,\frac{a(a+1)(a+2)b(b+1)(b+2)}{3!\,c(c+1)(c+2)}x^3 + \dots \\ &= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 \frac{t^{b-1}\,(1-t)^{c-b-1}}{(1-tx)^a}\,dt \end{aligned} \tag{1.1}$$

The integral on which Riemann applied the saddle-point method is given by

$$I_n = \int_0^1 s^{a+n}\,(1-s)^{b+n}\,(1-xs)^{c-n}\,ds = \int_0^1 \left(\frac{s\,(1-s)}{1-xs}\right)^n s^a\,(1-s)^b\,(1-xs)^c\,ds$$

where a, b, c are real, x is complex and n is large (Petrova & Solov'ev, 1997:368). The integral I_n is closely related to the regular solution (1.1). Riemann found that

$$I_n \sim \sqrt{\frac{\pi}{n}}\,\frac{\left(\sqrt{1-x}\right)^{b+c+\frac{1}{2}}}{\left(1-\sqrt{1-x}\right)^{2n+a+b+1}} \qquad \text{as} \quad n\to\infty$$

(Petrova & Solov'ev, 1997:369).

The Bessel Functions

Debeye (1909) uses the saddle-point method to derive asymptotic estimates of the Bessel functions. The Bessel functions are solutions to the Bessel differential equation

$$\frac{d^2y}{dx^2} + \frac{1}{x}\,\frac{dy}{dx} - \left(1 - \frac{a^2}{x^2}\right)y = 0$$

Debeye first focuses on estimating so-called Hankel functions (which are also called Bessel functions of the third kind) and then estimates other Bessel functions afterward. Hankel functions are two linearly independent solutions of Bessel's differential equation and are given by

$$H_1(x) = -\frac{1}{\pi}\int_{C_1} e^{-ix\,\sin z + iaz}\,dz \qquad \text{and} \qquad H_2(x) = -\frac{1}{\pi}\int_{C_2} e^{-ix\,\sin z + iaz}\,dz$$

where C_1 and C_2 are paths (or contours) illustrated by Figure 2 of Petrova & Solov'ev (1997:373). Debeye uses a saddle-point method together with Watson's lemma to derive asymptotic expansions for the two Hankel functions as $x\to\infty$ such that $\zeta = \frac{a}{x}$ is constant. He then proceeds to derive asymptotic expansions of the other Bessel functions by using the same method (Petrova & Solov'ev, 1997:374).

The Airy Functions

There are four types of Airy functions, denoted by $Ai(x)$, $Bi(x)$, $Gi(x)$ and $Hi(x)$. The Airy functions $Ai(x)$ and $Bi(x)$ are defined as the two linearly independent solutions to the Airy (or Stoke's) equation

$$\frac{d^2y}{dx^2} - y\,x = 0$$

The Airy function $Ai(x)$ is given by

$$Ai(x) = \frac{1}{\pi}\int_0^\infty cos\left(\frac{t^3}{3} + xt\right) dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ixs+\frac{1}{3}is^3} ds$$

The saddle-point method is applied to $Ai(x)$ to yield an asymptotic estimate as $x \to \infty$ (Lopez et al., 2009).

The Airy function $Ai(x)$ appears in optics, quantum mechanics and radiative transfer. In quantum mechanics, the Airy function $Ai(x)$ is a solution to Schrodinger's equation for a specific particle and is closely related to the so-called WKB approximation, which plays an important role in quantum field theory.

Cauchy Coefficient Integrals

Flajolet & Sedgewick (2009:554) use the saddle-point method to estimate Cauchy coefficient integrals of a generating function $G(z)$. If the generating function $G(z)$ is analytic around the origin on a disc $D(0, r)$ then $G(z)$ has the Taylor series:

$$G(z) = a_0 + a_1 z + a_2 z^2 + \ldots \qquad \text{where} \quad a_n = \frac{1}{2\pi i}\int_{C(0,r)} \frac{G(z)}{z^{n+1}}\, dz$$

The coefficients $a_0, a_1, \ldots$ are called the Cauchy coefficients.

As an example, Flajolet & Sedgewick (2009) use the saddle-point method to estimate the Cauchy coefficients of the exponential generating function $G(z) = e^z$. They find that

$$a_n = \frac{1}{2\pi i}\int_{C(0,r)} \sim \frac{e^n}{n^n\sqrt{2\pi\, n}} \qquad \text{as} \quad n \to \infty$$

(Flajolet & Sedgewick, 2009:555-557).

The Riemann Zeta Function

Riemann uses a saddle-point method to derive an asymptotic estimate of the function $Z(t)$ which is related to the Riemann zeta function $\zeta(z)$ in the following way:

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z} \qquad \text{and} \qquad Z(t) = e^{i\,\theta(t)}\,\zeta\left(\frac{1}{2} + it\right)$$

where z is a complex number and $\theta(t)$ is chosen so that $Z(t)$ is a real function. The functions $Z(t)$ and $\theta(t)$ are called the Riemann-Siegel functions.

The function $Z(t)$ can be written as

$$Z(t) = 2\sum_{n=1}^{N} \frac{\cos\theta(t) - t\,\log n}{\sqrt{n}} + R_N(t) \tag{1.2}$$

where

$$R_N(t) = -\frac{e^{i\,\theta(t)}\,\Gamma\left(\frac{1}{2} - it\right)}{2\pi\,i} \int_{C_N} \frac{(-z)^{-\frac{1}{2}+it}\,e^{-Nz}}{e^z - 1}\,dz$$

(Berry, 2016:72). The sum in (1.2) is called the "main sum" and $R_n(t)$ is called the "approximate functional equation". Both of these parts are real.

By choosing $N = \left\lfloor\sqrt{\frac{t}{2\pi}}\right\rfloor$ and defining $p = N - \sqrt{\frac{t}{2\pi}}$, Riemann proved that

$$R_N(t) = \left(\frac{2\pi}{t}\right)^{\frac{1}{4}} (-1)^{N+1}\,\frac{\cos\left(2\pi(p^2 - p - \frac{1}{16})\right)}{\cos 2\pi p} + O\left(t^{-\frac{1}{2}}\right)$$

(Berry, 2016:72-74).

1.4 The Saddle-point Approximation

1.4.1 Introduction

The saddle-point approximation historically refers to the uniform asymptotic expansion of the distribution of certain random variables which depend on a large sample size. However, due to its high accuracy and good behaviour in approximating the distribution of random variables which do not depend on large parameters, its use has become popular in the non-asymptotic setting. The saddle-point approximation presently refers to the saddle-point density and distribution functions as defined by Butler (2007). This saddle-point approximation can be applied to the distribution of any random variable with a well-behaved cumulant generating function.

This section has two parts. Firstly, a review of the saddle-point approximation is given in section 1.3.2. Secondly, some of the applications of the saddle-point approximation are discussed in section 1.3.3. The theory underlying the saddle-point approximation is discussed later in chapter 3.

1.4.2 A Review of the Saddle-point Approximation in Statistics

Daniels introduces the saddle-point approximation into statistics in 1954 with his "pioneering paper" (Barndorff-Nielsen & Cox, 1979:281) entitled *Saddlepoint Approximations in Statistics*. In this paper, Daniels applies the saddle-point method to the Fourier inversion formula of a density function of a sample mean to yield a uniform asymptotic expansion of said sample mean. He further proves that this saddle-point asymptotic expansion is an improvement of the Edgeworth expansion density function of a sample mean, which is itself an improvement of the normal approximation of a sample mean.

The insights of Daniels later provides a cornerstone for further applications of the saddle-point method in statistics, although it lay dormant for a few decades until the publication of the paper by Barndorff-Nielsen & Cox (1979). This publication is followed by numerous others which led to the establishment of the saddle-point method as an invaluable statistical method.

Barndorff-Nielsen & Cox (1979) uses exponential tilting together with the Edgeworth expansion to derive a uniform asymptotic expansion of the density function of a sample mean. This asymptotic expansion turns out to be equivalent to the saddle-point asymptotic expansion of Daniels (1954).

In the renowned paper by Lugannani & Rice (1980) they apply a saddle-point method similar to that of Bleistein (1966) to the Fourier inversion formula of a distribution function of a sample mean to yield a uniform asymptotic expansion of said sample mean. This uniform asymptotic expansion for the distribution function of a sample mean is popularly known as the Lugannani-Rice formula.

Daniels (1982) derives the saddle-point approximation for the probabilities of a time-homogeneous birth process. He achieves this by applying the saddle-point method on the Laplace inversion formula. He noted that in many cases the saddle-point approximation is exact and in general is very nearly exact.

In Barndorff-Nielsen (1983) he elaborates on his previous work with Cox (see Barndorff-Nielsen & Cox (1979)) to approximate the density function of a maximum likelihood estimator

and the likelihood ratio statistic.

In Daniels (1983) he derives two saddle-point approximations for the distribution of an estimator which is defined by an estimating equation. The second of these saddle-point approximations is a generalisation of the approximation of Field & Hampel (1982).

Easton & Ronchetti (1986) propose a general saddle-point approximation. This saddle-point approximation entails approximating the cumulant generating function by estimating the first four cumulants and then applying the saddle-point approximation to this approximated cumulant generating function. They illustrate the accuracy of the general saddle-point approximation by applying it to L-statistics.

Daniels (1987) discusses two methods of finding asymptotic expansions of the distribution function of a sample mean. The first method entails exponential tilting and the application of the Edgeworth expansion. The resulting asymptotic expansion is not uniform and is only accurate in the upper tail. The second method is the saddle-point method, otherwise known as the method of steepest descent. The saddle-point method yields the formula of Lugannani & Rice (1980) as the asymptotic expansion. This asymptotic expansion is uniformly accurate over the domain of the distribution. Daniels concludes that the two methods yield different approximations of the distribution function of a sample mean. This is contrary to the density function of the sample mean, in which case both methods yield the same approximations.

The discussion paper by Reid (1988) provides a "masterly review" of the saddle-point method in statistics according to Daniels (1988). The paper starts with a summary of the derivation of the saddle-point approximation and discusses both the saddle-point method and indirect Edgeworth expansion approach. Thereafter some of the applications of the saddle-point approximation are discussed. These applications include the maximum likelihood estimator, the likelihood ratio statistic and the score statistic.

Hougaard (1988), in his comments on the paper by Reid (1988), demonstrates how the saddle-point method can be applied to provide an accurate closed-form approximation of a non-closed form density function. He applies the saddle-point method to approximate the density function of a non-central gamma distribution and finds that the approximation provides an almost-perfect fit once it has been normalised. This application by Hougaard (1988) is the first application of the saddle-point approximation within a non-asymptotic setting.

The monograph by Field & Ronchetti (1990) discusses the saddle-point approximation for small samples as an alternative to the Edgeworth expansion. They furthermore discuss a general saddle-point approximation which is applied to M-estimators and L-estimators. This general saddle-point method entails applying the saddle-point approximation to an approximated cumulant generating function as is done by Easton & Ronchetti (1986).

The discussion paper by Goutis & Casella (1999) provides an elementary explanation of the two approaches of deriving the saddle-point approximation. They first discuss the saddle-point method and how it can be applied to yield the saddle-point approximation. Thereafter they discussed the approach of using exponential tilting together with Edgeworth expansions.

Butler's (2007) book is written to "provide an accessible account of the theory and applications of the saddle-point approximation that can be understood by the widest possible audience" (Butler, 2007). The book includes most of the saddle-point approximation's applications. Such

applications include the general formulas for the saddle-point density and distribution functions, the p^* and r^* formulas for maximum likelihood estimators, the saddle-point approximation for stochastic systems and saddle-point approximation in multivariate testing and the saddle-point approximation of estimators defined by estimating equations.

1.4.3 Applications of the Saddle-point Approximation

This subsection discusses some of the existing applications of the saddle-point approximation in statistics. The saddle-point approximation is highly accurate and can be applied to the distribution of any random variable which has a well-behaved cumulant generating function. A brief discussion of a few selected applications is provided, the purpose being to emphasise the variety of ways in which the saddle-point approximation can be applied.

Distribution of Maximum Likelihood Estimators

The following discussion of the saddle-point approximation of the distribution of maximum likelihood estimators follows closely that of Butler (2007). The p^* density is introduced by Barndorff-Nielsen (1980, 1983) as an approximation for the density of the maximum likelihood estimate, and is derived from the saddle-point approximation (Butler, 2007:219). The unnormalised saddle-point density of the maximum likelihood estimator $\hat{\underline{\theta}}$ is given by

$$p(\hat{\underline{\theta}};\underline{\theta}) = (2\pi)^{-\frac{m}{2}} \left| j(\hat{\underline{\theta}})^{\frac{1}{2}} \right| \frac{L(\underline{\theta})}{L(\hat{\underline{\theta}})}$$

where $j(\hat{\underline{\theta}})$ is the observed Fisher information and $\frac{L(\underline{\theta})}{L(\hat{\underline{\theta}})}$ is the likelihood ratio (Butler, 2007). Normalising the above density function with the constant $c(\underline{\theta})$ results in the p^* density function:

$$p^*(\hat{\underline{\theta}};\underline{\theta}) = c(\underline{\theta}) \left| j(\hat{\underline{\theta}})^{\frac{1}{2}} \right| \frac{L(\underline{\theta})}{L(\hat{\underline{\theta}})}$$

Barndorff-Nielsen further derives the r^* approximation for the continuous univariate cumulative distribution functions of maximum likelihood estimates. The r^* approximation is further discussed in chapter 8 of Butler (2007).

Weighted Sums of Random Variables

The saddle-point approximation given by Butler (2007) can be applied to approximate the distribution of a weighted sum of non-identical independent random variables. These random variables can be from different distributions. As an example in his book, Butler (2007) determines the distribution of the sum of a normal and a Laplacian random variable. The sum in this example can easily be extended to include more random variables, each with a corresponding weight.

Stochastic Processes

The saddle-point approximation can be used to approximate first-passage and time-to-event distributions (Butler, 2007). Gatto (2007, 2010) applies the saddle-point approximation in the compound Poisson process context to approximate, among other items, the probability of ruin and the discounted claim amounts. Renshaw (2000) applies the saddle-point approximation to bivariate stochastic processes.

Anderson-Darling Test Statistic

Giles (2006) derives the saddle-point approximation of the distribution of the Anderson-Darling test statistic and found that it "is markedly superior to other theoretical approximations in the lower tail of the distribution." Chen & Giles (2008) compare the various saddle-point approximations of Wood et al. (1993) to the saddle-point approximation of Lugannani and Rice (1980) and found that the former performs best for the Anderson-Darling test statistics. Furthermore, Murakami (2009) derives the saddle-point approximation of the modified Anderson-Darling test statistic and concluded that it is superior to other theoretical approximations.

Regression Estimators

Spady (1991) derives the saddle-point approximation of the distribution of regression estimators which are defined by a system of estimating equations where the observations are assumed to be independent. He further illustrates the accuracy of the saddle-point approximation by applying it to approximate the distribution of the L_1 regression estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

The Bootstrap

The Bootstrap method is a simulation-based method of approximating the distribution of a statistic. The non-parametric Bootstrap entails simulating samples from the empirical distribution function, determining the corresponding statistic for each sample and then approximating the distribution of the statistic using these "replicas". The saddle-point approach to performing the Bootstrap entails simulating samples from the empirical saddle-point distribution function, determining the corresponding statistic for each sample and then approximating the distribution of the statistic using these "replicas". The empirical saddle-point distribution function is the saddle-point distribution function which corresponds to the empirical cumulant generating function (Davison & Hinkley, 1988).

Option Pricing

Rogers and Zane (1999) use the saddle-point approximation to price European options. European option prices have historically been calculated using the fast Fourier transform (FFT) which is a numerical method. The saddle-point approximation has two advantages to a numerical method: it is not a recursive method and it is of an analytic form (Rogers & Zane, 1999:494).

In this subsection a selected number of applications of the saddle-point method are briefly discussed. Further applications are available in the literature.

The following chapter provides a broad overview of the saddle-point method. The saddle-point method is adapted and applied in the third chapter to yield the saddle-point approximation.

Chapter 2

Overview of the Saddle-point Method

2.1 Introduction

The saddle-point method is an approach to deriving an asymptotic estimate of a complex integral. The method is applied to the Fourier inversion formula of the probability density function and cumulative distribution function to yield the saddle-point density and distribution functions respectively. The utility of the saddle-point method lies in the flexibility of its methodology and conditions. The strictness of the conditions is reflected in the result which the saddle-point method yields - the stronger the conditions the more powerful the saddle-point method's result. Under certain strict conditions, the saddle-point method produces a uniform asymptotic expansion of both the probability density function and the cumulative distribution function. Under weak conditions, the saddle-point method produces accurate approximations. Although the saddle-point method yields an asymptotic result, the method can be adapted to provide an accurate approximation when there is no large parameter present. The saddle-point method has a flexible methodology in the sense that new saddle-point methodology can be developed to solve new problems. Two saddle-point methods are adapted to approximate the probability density function and cumulative distribution function respectively. These two methods are discussed in sections 2.3.5 and 2.3.6 and their adapted counterparts are discussed in sections 3.2.3 and 3.3.3.

The aim of this chapter is to provide an overview of the two saddle-point methods which are used in the next chapter. This chapter has two parts.

The mathematical theory underlying the saddle-point method is discussed first. An understanding of complex analysis, specifically complex integration, is necessary in order to appreciate the saddle-point method. The discussion of complex analysis covers the important concepts which relate to complex integration. These include paths, the differentiability of complex functions and the landscape formed by the modulus of a complex function.

Thereafter, the two saddle-point methods are discussed. The first saddle-point method is the method Flajolet & Sedgewick (2009:552), which is a generalisation of the method of de Bruijn (1958). The second method is a generalisation of the method of Bleistein (1966) which compensates for a simple pole at the origin. These two methods are adapted in chapter 3 to yield the saddle-point density and distribution functions respectively.

2.2 Complex Analysis

2.2.1 Introduction

The saddle-point method is based in complex analysis and is used to derive an asymptotic estimate of a complex integral (which is also called a contour integral). An understanding of complex analysis is essential to understanding the saddle-point method. This section is set out to cover the important topics in complex analysis that are relevant to the saddle-point method. These topics include paths, differentiability, analytic functions and the maximum modulus principle.

Complex analysis is the field in mathematics in which the complex space (denoted $\mathbb{C}$) is studied. Complex analysis originated from the study of algebraic equations with imaginary roots, such as $x^2 + 1 = 0$ (the roots being $\pm i$). The field has caught the interest of many well-known mathematicians. Euler produced and proved the formula $e^{i\theta} = \cos\theta + i\sin\theta$. Gauss used complex analysis to prove the Fundamental Theorem of the Algebra. Cauchy produced and proved several renowned theorems in complex analysis, including the Residue Theorem. The latter is used to solve difficult real integrals and find the limits of some real series which are unsolvable in real analysis.

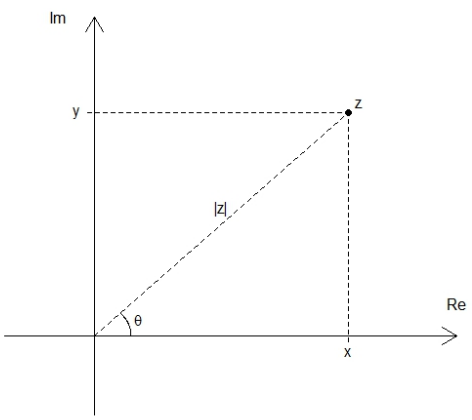
This section serves as an introduction to complex analysis with focus on complex integration. There are five parts to this section. Complex numbers and functions are discussed in sections 2.2.2 and 2.2.3. Section 2.2.4 concerns the differentiation of complex functions. The modulus of a complex function is discussed in section 2.2.5. Section 2.2.6 deals with complex integration and its properties.

2.2.2 Complex Numbers

A complex number z can be written as $z = x + iy = |z|e^{i\theta}$ where

$$
\begin{aligned}
&x = Re\{z\} \in \mathbb{R} \quad \text{is the real part of } z \\
&y = Im\{z\} \in \mathbb{R} \quad \text{is the imaginary part of } z \\
&i = \sqrt{-1} \quad \text{is the imaginary number} \\
&|z| = \sqrt{x^2 + y^2} \quad \text{is the modulus of } z \\
&\theta \quad \text{is called the argument of } z
\end{aligned}
$$

A complex number $z = x + iy = |z|e^{i\theta}$ can be illustrated by an Argand diagram, which is a Cartesian plot of z on the complex plane. An Argand diagram of z is given as follows:



The Argand diagram has two axes: the horizontal real axis and the vertical imaginary axis. The complex space is defined as the set of all complex numbers $\{x + iy : x, y \in \mathbb{R}\}$ and is denoted by $\mathbb{C}$.

2.2.3 Complex Functions and Paths

There are two types of complex functions:

A Complex Function of a Real Variable

A complex function of a real variable, for example $\phi : \mathbb{R} \to \mathbb{C}$, is of the form $\phi(s) = u(s) + i\, v(s)$ where $s \in \mathbb{R}$ and $u, v : \mathbb{R} \to \mathbb{R}$. Such a function can be plotted on the complex plane.

An important subset of such functions are called paths. A path is a continuous function $\phi : [s_1, s_2] \to \mathbb{C}$ restricted to $[s_1, s_2] \subset \mathbb{R}$ (Conway, 1973:45). Paths play an essential role in complex integration, which is discussed in the section 2.2.6.

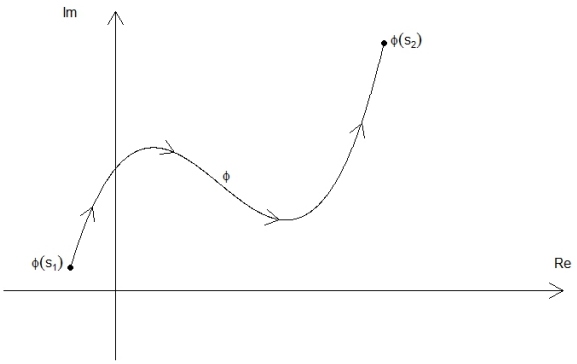


Figure 2.1: A path ϕ plotted on the complex plane. The direction of the path is indicated by the arrows.

Some of the necessary terminology associated with paths is given in order to make it easier to discuss paths and complex integration. The range of a path ϕ, defined as $\phi^* = \{\phi(s) : s \in [s_1, s_2]\}$, is the set of all the complex points that lie on the path ϕ. The backward path of ϕ is given by $-\phi$ where $(-\phi)(s) = \phi(-s)$ for $s \in [-s_2, -s_1]$. $\phi(s_1)$ and $\phi(s_2)$ are called the start-point and end-point of the path ϕ respectively. The length of a path is defined as $L(\phi) = \int_{s_1}^{s_2} \phi^{(1)}(s)\, ds$. The derivative of a complex function $\phi^{(1)}(s)$ is discussed in the following section. A path is closed if its start-point and end-point are the same (i.e. $\phi(s_1) = \phi(s_2)$).

A Complex Function of a Complex Variable

A complex function of a complex variable, for example $g : \mathbb{C} \to \mathbb{C}$, is of the form $g(x + iy) = u(x, y) + i\, v(x, y)$ where $x, y \in \mathbb{R}$ and $u, v : \mathbb{R}^2 \to \mathbb{R}$. Such a function cannot be plotted on the complex plane.

The modulus of such a function is defined as $|g(x + iy)| = \sqrt{u^2(x, y) + v^2(x, y)}$. The modulus $|g(z)|$ can be plotted as a three dimensional graph as a function of x and y, and can be viewed as a landscape. The modulus of an analytic function plays a pivotal role in the saddle-point method which is discussed in section 2.3.5.

2.2.4 Complex Differentiation and Analytic Functions

The function $g : \mathbb{C} \to \mathbb{C}$, which can be written as $g(x+iy) = u(x,y) + i\,v(x,y)$, is differentiable at $z_0 = x_0 + iy_0$ if and only if

- u and v are differentiable on $\mathbb{R}^2$
- $$\left.\frac{\partial u(x,y)}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} = \left.\frac{\partial v(x,y)}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} \quad \text{and} \quad \left.\frac{\partial u(x,y)}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} = -\left.\frac{\partial v(x,y)}{\partial x}\right|_{\substack{x=x_0\\y=y_0}}$$

The two equations given under the second bullet-point are called the Cauchy-Riemann equations. If the two conditions hold then

$$\begin{aligned} g^{(1)}(x_0+iy_0) &= \left.\frac{\partial u(x,y)}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} + i\left.\frac{\partial v(x,y)}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} \\ &= \left.\frac{\partial v(x,y)}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} - i\left.\frac{\partial u(x,y)}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} \end{aligned}$$

Analytic Functions

A complex function $g : \mathbb{C} \to \mathbb{C}$ is said to be analytic on an open set $\Omega \subset \mathbb{C}$ if it is infinitely differentiable on the set Ω. In layman's terms, an analytic function is well-behaved and smooth since it can be differentiated ad infinitum. For the saddle-point method, it is assumed that the function being integrated is analytic.

Singularities

The singularities of a complex function are the set of complex points $z_1, z_2, \ldots$ where the function is not analytic. For instance, the function $g(z) = \frac{1}{1-z}$ has a singularity at $z = 1$. Singularities play an important role in the saddle-point method discussed in section 2.4.5.

2.2.5 The Modulus of an Analytic Function

The Modulus of a Function

The modulus of a function $g(x+iy) = u(x,y) + i\,v(x,y)$ is defined in section 2.2.3 as

$$|g(z+iy)| = \sqrt{u^2(x,y) + v^2(x,y)}$$

It is possible to write the function as $g(x,y) = |g(x,y)|\, e^{i\,\theta(x,y)}$ where $\theta(x,y)$ is the argument of $g(x+iy)$. It follows that if $h(z) = \log g(z)$ then $|g(z)| = e^{Re\{h(z)\}}$ where $Re\{h(z)\}$ is the real part of $h(z)$.

In sections 2.3.4 and 2.3.5 the terms $Re\{h(z)\}$ and $|g(z)|$ are used interchangeably since there exists a one-to-one relationship between them, namely $Re\{h(z)\} = \log|g(z)|$.

The Maximum Modulus Principle

Consider a function g which is analytic over $\Omega \subset \mathbb{C}$. If there exists a point $z_0 \in \Omega$ such that $|g(z_0)| \geq |g(z)|$ for all z in a neighbourhood of z_0, then g is constant (Conway, 1973:124).

In other words, if there exists a point $z_0 \in \Omega$ at which the modulus $|g|$ has a local maximum, then g is constant.

The Surface Generated by $|g|$

This subsection follows closely that of Flajolet & Sedgewick (2009:543). If g is analytic on $\Omega \subset \mathbb{C}$ and not-constant then $|g|$ has no local maxima on Ω. The surface generated by the modulus $|g|$ has points of only three possible types: zero points, ordinary points and so-called saddle-points (Flajolet & Sedgewick, 2009:543).

Zero points are defined as the set $\{z \in \Omega : g(z) = 0\}$. The surface generated by $|g|$ attains its minimum value 0 at the zero-points. Flajolet & Sedgewick (2009) use the analogy that a zero point is like "the bottom of a lake, save that, in the landscape of an analytic function, all lakes are at sea level" (Flajolet & Sedgewick, 2009:543).

Ordinary points are defined as the set $\left\{z \in \Omega : g(z) \neq 0, g^{(1)}(z) \neq 0\right\}$. Most of the points on the surface generated by $|g|$ are ordinary points (Flajolet & Sedgewick, 2009:543)

Saddle-points are defined as the set $\left\{z \in \Omega : g(z) \neq 0, g^{(1)}(z) = 0\right\}$. Although it is not clear from the definition why such points are called "saddle-points", it can be shown that at a saddle-point the gradient of $|g|$ increases in one direction and decreases in another direction. This property of a saddle-point - that it is a minimum in one direction and a maximum in another - plays a central role in the saddle-point method. The surface generated by the $|g|$ forms the shape of a saddle around a saddle-point. This is illustrated in Figure 2.2.

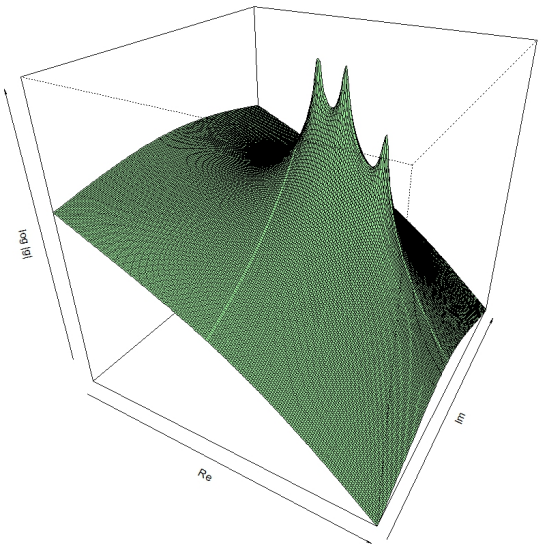


Figure 2.2: An illustration of a landscape $\log|g(z)|$ where $g(z) = \frac{1}{1-\frac{1}{2}z}\frac{1}{1-z}\frac{1}{1-2z}$. The function $g(z)$ has singularities at $\frac{1}{2}, 1$ and 2. These singularities correspond to the unending peaks on the illustration.

The landscape $|g|$ is often described as a mountain range, where the peaks of the mountains are too high to be seen (de Bruijn, 1958:82). As a mountaineer on this mountain range, it is possible to stand at three types of places: at the base of the mountain range (zero points), on the mountain range at a gradient (normal points) and on the highest point of a mountain pass, where the gradient is flat (a saddle-point).

This mountain range landscape of $|g|$ plays an essential role in determining the saddle-point path, which is discussed in section 2.3.4.

2.2.6 Complex Integration

The integral of a piece-wise continous function $g : \mathbb{C} \to \mathbb{C}$ over a piece-wise smooth path $\phi : [s_1, s_2] \to \mathbb{C}$ is defined as

$$\int_{\phi} g(z)\, dz = \int_{s_1}^{s_2} g\left(\phi(t)\right)\, \phi^{(1)}(t)\, dt$$

A complex integral depends on the function g on the choice of path ϕ. There are similarities between complex integration and real integration. For example, for a piece-wise smooth path from z_1 to z_2, the integral of g over this path is in fact the difference of the anti-derivative of g at the points z_2 and z_1 respectively. There are, however, many notable differences. One of the most notable differences is that in complex integration, the path of integration can be chosen (given certain conditions). Another notable difference is the landscape formed by the modulus of an analytic function. If the function is non-constant then it has no local maxima due to the maximum modulus principle.

The properties of complex integration make it useful in solving a variety of problems. One such problem-solving method is the saddle-point method. The properties of complex integration that play a central role in the derivation of the saddle-point method are discussed below.

Properties of Complex Integration

1 If the modulus $|g(z)|$ is less or equal to $M(\phi, g)$ for every point z on the path ϕ and $L(\phi)$ is the length of the path, then

$$\left| \int_{\phi} f(z)\, dz \right| \leq M(\phi, g)\, L(\phi)$$

2 If the paths ϕ_1 and ϕ_2 have the same start- and end-points and there are no singularities between them, then

$$\int_{\phi_1} f(z)\, dz = \int_{\phi_2} g(z)\, dz$$

In other words, it is possible to choose the path of integration, as long as the start- and end-points of the path remain the same.

3 If the paths ϕ_1 and ϕ_2 are piece-wise smooth such that the combined path $\phi_1 \cup \phi_2$ is also piece-wise smooth, then

$$\int_{\phi_1 \cup \phi_2} g(z)\, dz = \int_{\phi_1} g(z)\, dz + \int_{\phi_2} g(z)\, dz$$

This result is used in the derivation of the saddle-point method in section 2.3.5.

4 **Cauchy's Theorem**: If the function g is differentiable over an open set $\Omega \subset \mathbb{C}$

$$\int_{\phi} g(z)\,dz = 0$$

for any closed piece-wise smooth path ϕ in Ω.

5 If a transformation $w = w(z)$ is surjective, monotone increasing and continuous, then

$$\int_{\phi} g(z)\,dz = \int_{\phi} g(z(w))\,\frac{dz(w)}{dw}\,dw$$

This result is used in the derivation of the saddle-point method when there is a singularity present at the origin. This is discussed in section 2.3.6.

2.2.7 Remarks

The five properties of complex integration mentioned above are central to the discussion of the saddle-point method in section 2.3. The combination of properties 1 and 2 implies that there exist paths that minimise the upper bound of property 1. These paths are discussed in section 2.3.4 and are called the saddle-point paths. The combination of properties 3 and 4 implies that

$$\int_{a}^{b} g(z)\,dz = \int_{a}^{a_1} g(z)\,dz + \int_{a_1}^{b_1} g(z)\,dz + \int_{b_1}^{b} g(z)\,dz$$

This result is used in section 2.3.5 to split the integral $\int_a^b g(z)\,dz$ into the three parts above. Property 5 is used in section 2.3.6 in the discussion of the saddle-point method which compensates for a singularity at 0.

Analyticity and singularities have important implications in the derivation of the saddle-point method. For the basic saddle-point method, it is assumed that the function is analytic over all possible values of the saddle-point. Bleistein (1966), however, derives a saddle-point method which compensates for a singularity at the origin.

2.3 The Saddle-point Method

2.3.1 Introduction

The saddle-point method is an approach to derive an asymptotic estimate of a complex integral

$$I = \frac{1}{2\pi i} \int_{\phi} g(z)\, dz$$

where the complex function g is analytic on an open set $\Omega \subset \mathbb{C}$ and ϕ is a piece-wise smooth path from a to b in Ω. Furthermore, it is assumed that the complex function $g \equiv g_\lambda$ depends on some large parameter λ.

The saddle-point method can be regarded as the extension of Laplace's method in the complex space. de Bruijn (1958) derives the saddle-point method in two stages: the saddle-point path (which includes a discussion of peakless landscapes and steepest descent) and Laplace's method. Flajolet & Sedgewick (2009) also regard the saddle-point method as having two parts: the choice of path and Laplace's method

In this chapter, the saddle-point method is discussed in a manner similar to the expositions of de Bruijn (1958) and Flajolet & Sedgewick (2009). Firstly, Laplace's method is discussed. Secondly, the saddle-point path is discussed by relying on de Bruijn's analogy of the mountaineer. Thirdly, the saddle-point methods of de Bruijn (1958) and Flajolet & Sedgewick (2009) are discussed. An adaptation of the former method is used in section 3.2 to derive the saddle-point density function. Fourthly, the saddle-point method of Bleistein (1966) and a generalisation thereof are discussed. An adaptation of the generalised method of Bleistein (1966) is used in section 3.3 to derive the saddle-point distribution function.

2.3.2 Laplace's Method

Laplace's method is an approach to approximate a real integral of the form

$$I_\lambda = \int_{-\infty}^{\infty} e^{\lambda\, h(x)}\, dx$$

where $h(x)$ is a real and continuous function and λ is large. The discussion in this section follows closely that of de Bruijn (1958:60-68). Define the global maximum of $h(x)$ as $\hat{t}$ such that it satisfies the equations

$$h^{(1)}(\hat{t}) = 0 \qquad \text{and} \qquad h^{(2)}(\hat{t}) < 0$$

It is further required that there exist real positive numbers c and r such that

$$h(x + \hat{t}) \leq h(\hat{t}) - c \qquad \text{for all } |x| > r$$

Define the function $h^\star(x)$ as

$$h^\star(x) = h(x + \hat{t}) - h(\hat{t})$$

It follows from the definition of $h^\star(t)$ that it has a global maximum at 0 and $h(0) = 0$. For large λ, the function $e^{\lambda\, h^\star(x)}$ has a sharp peak at 0. Due to this sharp peak, the integral over some interval containing 0, say $(-\delta, \delta)$, is almost equal to the whole integral when λ is large. Mathematically speaking, we have the following three results:

1. The tail integrals of $h^\star(x)$ are negligible compared to the integral of $h^\star(x)$ over the interval $(-\delta, \delta)$ when λ is large:

$$\int_{-\infty}^{-\delta} e^{\lambda\, h^\star(x)} + \int_{\delta}^{\infty} e^{\lambda\, h^\star(x)} = o\left(\int_{-\delta}^{\delta} e^{\lambda\, h^\star(x)}\right) \qquad \text{as } \lambda \to \infty$$

2. The quadratic expansion $h^\star(x + \hat{t}) = \frac{1}{2}\, h^{(2)}(\hat{t})\, x^2 + O(x^3)$ is valid for all $x \in (-\delta, \delta)$.

3. The incomplete Gaussian integral over the interval $(-\delta, \delta)$ is asymptotically equivalent to the complete Gaussian integral as $\lambda \to \infty$.

It follows from de Bruijn (1985:63-65) that

$$I_\lambda^\star = \int_{-\infty}^{\infty} e^{\lambda\, h^\star(x)} dx \sim \int_{-\delta}^{\delta} e^{\lambda\, h^\star(x)} dx \sim \int_{-\delta}^{\delta} e^{\lambda\, h^{(2)}(\hat{t})\, x^2} dx = \frac{\sqrt{2\pi}}{\sqrt{-\lambda\, h^{(2)}(\hat{t})}} \qquad \text{as } \lambda \to \infty$$

de Bruijn (1958:66-69) further proves that $I_\lambda^\star$ has the asymptotic expansion

$$I_\lambda^\star = \int_{-\infty}^{\infty} e^{\lambda\, h^\star(x)} dx = \sum_{j=0}^{\infty} \frac{c_j}{\lambda^{j+\frac{1}{2}}} \qquad \text{as } \lambda \to \infty$$

for some real constants $c_0, c_1, c_2 \ldots$ where $c_0 = \frac{\sqrt{2\pi}}{\sqrt{-h^{(2)}(\hat{t})}}$.

By returning to the original integral I_λ, it follows that

$$I_\lambda = e^{\lambda\, h(\hat{t})}\, I_\lambda^\star = e^{\lambda\, h(\hat{t})} \sum_{j=0}^{\infty} \frac{c_j}{\lambda^{j+\frac{1}{2}}} \qquad \text{as } \lambda \to \infty \tag{2.1}$$

for some real constants $c_0, c_1, c_2 \ldots$ where $c_0 = \frac{\sqrt{2\pi}}{\sqrt{-h^{(2)}(\hat{t})}}$. This approximation holds under weak conditions and is easy to apply. As an example, Laplace's method is applied to the gamma function in the following subsection.

2.3.3 Example: Stirling's Formula

Consider the gamma function

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x}\, dx = \int_0^\infty e^{n\left(-\frac{x}{n} + \log x\right)}\, dx = n \int_0^\infty e^{n(-y + \log y + \log n)}\, dy = n^{n+1} \int_0^\infty e^{n\, h(y)}\, dy$$

where $h(y) = -y + \log y$ has a global maximum $\hat{t} = 1$. The further condition is satisfied by choosing $c = .1$ and any $r > .52$. Furthermore,

$$h(\hat{t}) = -\hat{t} + \log \hat{t} = -1 \qquad \text{and} \qquad h^{(2)}(\hat{t}) = -\frac{1}{\hat{t}^2} = -1$$

It follows that Laplace's method yields

$$\Gamma(n+1) \sim n^{n+1}\, \frac{\sqrt{2\pi}\, e^{n\, h(\hat{t})}}{\sqrt{-n\, h^{(2)}(\hat{t})}} = \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n} \qquad \text{as } n \to \infty$$

which is called Stirling's approximation.

2.3.4 The Saddle-point Path

Consider the complex integral

$$I = \frac{1}{2\pi i}\int_{\phi} g(z)\,dz = \frac{1}{2\pi i}\int_{\phi} e^{h(z)}\,dz \tag{2.2}$$

where the complex function $g(z)$ is analytic on an open set $\Omega \subset \mathbb{C}$ and $\phi(s)$ is a piece-wise smooth path from a to b in Ω.

The first property of section 2.2.6 states that the modulus of the integral is bounded as follows:

$$|I| \leq M(\phi, g)\, L(\phi)$$

The second property of section 2.2.6 implies that the path of integration ϕ can be chosen arbitrarily on condition that it starts at a and ends at b and that there are no singularities between the original path and the chosen path.

These two properties imply that there exists paths (from a to b) which minimise the upper bound $M(\phi, g)\, L(\phi)$. The length of the path $L(\phi)$ "usually turns out to be unimportant in asymptotic bounding purposes" (Flajolet & Sedgewick, 2009:547) and "is, as a rule, quite unimportant" (de Bruijn, 1958:78). The focus is therefore on determining the paths that minimise $M(\phi, g)$ - an upper bound of $|g|$ on the path ϕ. These paths are called the set of saddle-point paths of g and are defined as

$$\mathbb{P}(g) = \arg\inf_{\phi} M(\phi, g) = \arg\inf_{\phi}\sup_{z\in\phi^*} |g(z)| \tag{2.3}$$

The reason for the name "saddle-point path" is discussed under the following heading.

Hiking over the Modulus Landscape

A saddle-point of g is defined in section 2.2.5 as the solution $\hat{t} \in \mathbb{C}$ to the following two equations:

$$g(\hat{t}) \neq 0 \qquad \text{and} \qquad g^{(1)}(\hat{t}) = 0$$

As discussed in section 2.2.5. the surface generated by the modulus $|g(z)|$ can be regarded as a mountain range where the mountain peaks are too high to be seen, and the saddle-points of g can be regarded as mountain passes. If a and b lie on opposite sides of a mountain pass (i.e. on either side of a saddle-point), then the minimisation problem (2.3) is solved by the set of paths from a to b that crosses the mountain pass through the saddle-point.

This analogy is borrowed from de Bruijn (1958:80) who says that "if there exists a path that solves the minimum problem, the highest point of the path will be a saddle-point, that is, in his terminology, the highest point of a pass." (de Bruijn, 1958:80) referring to (2.3).

Similarly, Flajolet & Sedgewick (2009:548) in referring to (2.3) state that "if a and b lie in opposite valleys of a saddle-point, then the minimisation problem is solved by saddle-point paths made of arcs connecting a and b through the saddle-point."

The upper bound $M(\phi, g)\, L(\phi)$ is minimised by the saddle-point path which has the shortest length $L(\phi)$. In deriving the saddle-point method, however, we are interested in the steepest saddle-point path. This is discussed under the following heading.

Steepest Descent and the Saddle-point Axis

Laplace's method is most effective when the function being integrated has a sharp peak. Similarly, the saddle-point method is most effective when the path of integration is a steep saddle-point path. Such a path is called a path of steepest descent.

de Bruijn defines the saddle-point axis as the straight line α which passes through the saddle-point $\hat{t}$ and has the argument

$$\arg \alpha = \frac{\pi - \arg\{h^{(2)}(\hat{t})\}}{2} \tag{2.4}$$

At the saddle-point $\hat{t}$, the decrease in $|g|$ is strongest in the direction of the axis α. The directions of the saddle-point axis are therefore called the directions of steepest descent.

From this result, de Bruijn defines a path of steepest descent as a path from a to b which is chosen so that "its tangent at the saddle-point coincides with the saddle-point axis." If a mountaineer were following a path of steepest descent, then he would approach the mountain pass so that the path in which he crossed over it would be the steepest.

The analogy of de Bruijn (1985:80) again illustrates the path of steepest descent from the view of a mountaineer: "Let us imagine a man who wants to move from a to b in some mountain district, and whose physical condition makes it desirable to avoid the higher altitudes as much as possible. On the other hand, he has no objection whatsoever against walking, nor against climbing. He therefore tries to do the same thing as we want to do on our surface $|g|$: he wants to take the path such that the maximum altitude is as low as possible." Furthermore, due to the hiker's problem with altitude, his optimal path would also be path of steepest descent.

The use of steepest descent is not "strictly essential for the saddle-point method" (de Bruijn, 1985:85) as long as the path of integration does not deviate from the direction of steepest descent with an angle of more than $\frac{\pi}{4}$ (or 45 degrees).

2.3.5 The Saddle-point Method

An Extension of Laplace's Method to the Complex Space

First consider a complex integral of the form

$$I_\lambda = \frac{1}{2\pi i} \int_a^b e^{\lambda\, h(z)}\, dz \tag{2.5}$$

where the complex function $h(z)$ is analytic on an open set $\Omega \subset \mathbb{C}$, the path of integration is a piece-wise smooth path from a to b in Ω and λ is large. Note that in this and the following sections $h(z)$ is a complex function of a complex variable. As discussed in section 2.2.5, a saddle-point of $g(z) = e^{h(z)}$, denoted $\hat{t}$, is a solution to the following equations

$$g(\hat{t}) \neq 0 \text{ and } g^{(1)}(\hat{t}) = 0 \qquad \text{or equivalently} \qquad h^{(1)}(\hat{t}) = 0$$

The modulus can be written as $\left|e^{h(z)}\right| = e^{Re\{h(z)\}}$. The terms $Re\{h(z)\}$ and $|g(z)|$ are used interchangeably since there exists a one-to-one relationship between them, namely $Re\{h(z)\} = \log|g(z)|$.

The following exposition follows closely that of de Bruijn (1958:88). The integration path from a to b can be deformed to become a steepest-descent path joining the points a to a_1 to b_1 to b where a_1 and b_1 are points on the saddle-point axis with one on either side of the saddle-point $\hat{t}$.

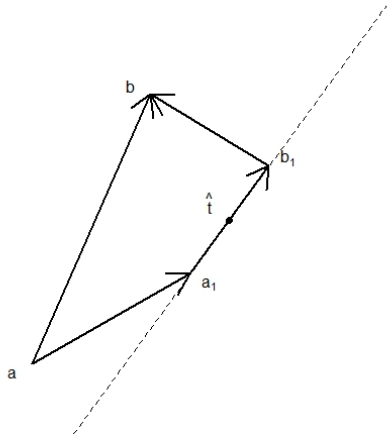


Figure 2.3: The dashed line represents the saddle-point axis.

The points a_1 and b_1 are chosen so that the sharp peak of $e^{\lambda\, Re\{h(z)\}}$ lies on the straight path between them. The choice of a_1 and b_1 implies that any point z on either of the paths from a to a_1 and b_1 to b has a log relative height $Re\{h(z)\} - Re\{h(\hat{t})\} < c < 0$ for some negative constant c. This implies that the tail integrals have a contribution of $O\left(e^{\lambda\, c}\, e^{\lambda\, Re\{h(\hat{t})\}}\right)$ (de Bruijn, 1985:88). Therefore

$$\begin{aligned}\int_a^b e^{\lambda\, h(z)}\, dz &= \int_a^{a_1} e^{\lambda\, h(z)}\, dz + \int_{a_1}^{b_1} e^{\lambda\, h(z)}\, dz + \int_{b_1}^{b} e^{\lambda\, h(z)}\, dz \\ &= \int_{a_1}^{b_1} e^{\lambda\, h(z)}\, dz + O\left(e^{\lambda\, c}\, e^{\lambda\, Re\{h(\hat{t})\}}\right) \\ &\sim \int_{a_1}^{b_1} e^{\lambda\, h(z)}\, dz \qquad \text{as } \lambda \to \infty \end{aligned} \tag{2.6}$$

Consider the substitution $z = \hat{t} + x\, e^{i\,\frac{\pi-\theta}{2}} = \hat{t} + i\, x\, e^{-i\,\frac{\theta}{2}}$ where θ is the argument of $h^{(2)}(\hat{t})$ and $\frac{\pi-\theta}{2}$ is the argument of the saddle-point axis. It follows from de Bruijn (1958:88) that

$$\int_{a_1}^{b_1} e^{\lambda\, h(z)}\, dz = i\, e^{-i\,\frac{\theta}{2}} \int_{a_2}^{b_2} e^{\lambda\, h\left(\hat{t}+i\, x\, e^{-i\,\frac{\theta}{2}}\right)}\, dx \qquad \text{for some real numbers } a_2 < 0 < b_2 \tag{2.7}$$

to which Laplace's method can be applied. The quadratic expansion of $h(\hat{t} + i\, x\, e^{-i\frac{\theta}{2}})$ around $\hat{t}$ is given by $h(\hat{t} + i\, x\, e^{-i\frac{\theta}{2}}) = h(\hat{t}) - \frac{1}{2}h^{(2)}(\hat{t})\, x^2\, e^{-i\theta} + \cdots = h(\hat{t}) - \frac{1}{2}\left|h^{(2)}(\hat{t})\right| x^2 + \ldots$. It follows that the application of Laplace's method to equation (2.7) yields the asymptotic expansion

$$\int_{a_1}^{b_1} e^{\lambda\, h(z)}\, dz = i\, e^{-i\,\frac{\theta}{2}}\, e^{\lambda\, h(\hat{t})} \sum_{j=0}^{\infty} \frac{c_j}{\lambda^{j+\frac{1}{2}}} \qquad \text{as } \lambda \to \infty \tag{2.8}$$

for some real constants $c_0, c_1, c_2 \ldots$ where $c_0 = \frac{\sqrt{2\pi}}{\sqrt{\lambda\,|h^{(2)}(\hat{t})|}}$. This is a complete asymptotic expansion in terms of λ.

Finally, by combining equations (2.6) and (2.8), it follows that

$$I_\lambda \sim \frac{1}{2\pi i}\int_{a_1}^{b_1} e^{\lambda\, h(z)}\, dz = \frac{e^{\lambda\, h(\hat{t})}}{\sqrt{2\pi\, \lambda\, h^{(2)}(\hat{t})}}\left(1+\frac{c_1}{\lambda}+\frac{c_2}{\lambda^2}+\ldots\right) \qquad \text{as } \lambda\to\infty \tag{2.9}$$

This method is an extension of Laplace's method in complex analysis. The form of the integral I_λ is Laplacian while the function h changes from real to complex.

The Fourier inversion formula of the probability density function of a sample mean has the form of (2.5). Result (2.9) can therefore be applied to yield an asymptotic expansion of the density function for a sample mean.

The Fourier inversion formulae of other probability density functions, however, are often not of the form of (2.5). It is therefore essential to derive a saddle-point method for a more general integral. The saddle-point method for this general integral, in turn, has a wider variety of applications than does (2.9).

A More General Integral

Consider a complex integral of the form

$$I_\lambda = \frac{1}{2\pi i}\int_a^b g_\lambda(z)\, dz = \frac{1}{2\pi i}\int_a^b e^{h_\lambda(z)}\, dz \tag{2.10}$$

where $g_\lambda(z) = e^{h_\lambda(z)}$ is analytic on an open set $\Omega \subset \mathbb{C}$, the path of integration is a piece-wise smooth path from a to b in Ω and the parameter λ is large.

Again, the integration path from a to b can be deformed to become a steepest-descent path joining the points a to a_1 to b_1 to b where a_1 and b_1 are points on the saddle-point axis with one on either side of the saddle-point $\hat{t}$. The points a_1 and b_1 are chosen so that the sharp peak of $|g_\lambda(z)| = e^{Re\{h_\lambda(z)\}}$ lies on the straight path between them.

Flajolet & Sedgewick (2009:552) impose the following conditions:

1. The tail integrals are negligible in comparison to the central integral:

$$\int_a^{a_1} e^{h_\lambda(z)}\, dz + \int_{b_1}^{b} e^{h_\lambda(z)}\, dz = o\left(\int_{a_1}^{b_1} e^{h_\lambda(z)}\, dz\right) \qquad \text{as } \lambda\to\infty$$

2. Along the central path (from a_1 to b_1), the quadratic expansion

$$h_\lambda(z) = h_\lambda(\hat{t}) + \frac{1}{2}\, h_\lambda^{(2)}(\hat{t})\, \left(z-\hat{t}\right)^2 + O(\epsilon_\lambda)$$

is valid, with $\epsilon_\lambda \to 0$ as $\lambda\to\infty$ uniformly for z on the straight line between a_1 and b_1.

3. The incomplete Gaussian integral over the real central path is asymptotically equivalent to the complete Gaussian integral:

$$\int_{a_2}^{b_2} e^{\frac{1}{2}\left|h_\lambda^{(2)}(\hat{t})\right| x^2}\, dx \sim \int_{-\infty}^{\infty} e^{\frac{1}{2}\left|h_\lambda^{(2)}(\hat{t})\right| x^2}\, dx = \sqrt{\frac{2\pi}{\left|h_\lambda^{(2)}(\hat{t})\right|}} \qquad \text{as } \lambda\to\infty$$

If all the conditions of Flajolet & Sedgewick (2009) are satisfied then

$$\begin{aligned}
I_\lambda &= \frac{1}{2\pi i}\int_a^b e^{h_\lambda(z)}\,dz \\
&\sim \frac{1}{2\pi i}\int_{a_1}^{b_1} e^{h_\lambda(z)}\,dz && \text{by condition 1} \\
&= \frac{1}{2\pi}\,e^{-i\,\frac{\theta}{2}}\int_{a_2}^{b_2} e^{h_\lambda\left(\hat{t}+i\,x\,e^{-i\,\frac{\theta}{2}}\right)}\,dx \\
&\sim \frac{1}{2\pi}\,e^{-i\,\frac{\theta}{2}}\,e^{h_\lambda(\hat{t})}\int_{a_2}^{b_2} e^{-\frac{1}{2}\left|h_\lambda^{(2)}(\hat{t})\right|x^2}\,dx && \text{by condition 2} \\
&\sim \frac{e^{-i\,\frac{\theta}{2}}\,e^{h_\lambda(\hat{t})}}{\sqrt{2\pi\,\left|h_\lambda^{(2)}(\hat{t})\right|}} && \text{by condition 3} \\
&= \frac{e^{h_\lambda(\hat{t})}}{\sqrt{2\pi\,h_\lambda^{(2)}(\hat{t})}} \qquad \text{as } \lambda\to\infty
\end{aligned} \tag{2.11}$$

which they call the basic saddle-point method. This method does not provide a complete asymptotic expansion nor does it provide the order by which I_λ converges. It is necessary to know $h_\lambda(z)$ to determine the asymptotic expansion of I_λ. This method does, however, provide the asymptotic limit of I_λ as $\lambda\to\infty$ under the conditions stipulated by Flajolet & Sedgewick (2009).

The Fourier inversion formula of a probability density function can be written in the form of equation (2.10). If the above conditions are satisfied then the result of equation (2.11) can be applied to yield a asymptotic limit of the probability density function.

The Fourier inversion formula of a cumulative distribution function can be written in a form similar to equation (2.10) but with a slight alteration - it has a singularity at 0. It is necessary to adapt the saddle-point method to compensate for the singularity at 0 in order to apply the saddle-point method to this Fourier inversion formula. Such a method is devised by Bleistein (1966) and is discussed in the subsection that follows.

2.3.6 The Saddle-point Method and a Singularity at the Origin

Consider the integral

$$I_\lambda = \frac{1}{2\pi i}\int_a^b e^{\lambda\,h(z)}\,\frac{dz}{z^k} \tag{2.12}$$

for any $k = 1, 2, \ldots$ where the complex function $h(z)$ is analytic on an open set $\Omega\subset\mathbb{C}$, the path of integration is a piece-wise smooth path from a to b in $\Omega\backslash\{0\}$ and the parameter λ is large.

Let $\hat{t}$ denote the saddle-point of $e^{h(z)}$. The function $e^{\lambda\,h(z)}\frac{1}{z^k}$ has a singularity at $z=0$. If $\hat{t}$ is close to 0 then $e^{\lambda\,h(z)}\frac{1}{z^k}$ can have multiple saddle-points close to 0. The question then becomes: which saddle-point should be used? The lowest saddle-point or a combination of saddle-points? By returning to de Bruijn's analogy of the hiker who wants to "take a path such that the maximum altitude is as low as possible", it is clear that the lowest saddle-points should be chosen. However, two or more saddle-points could be at roughly the same height. Furthermore, how much lower should a saddle-point be (relative to the other saddle-points)

for it to lead to an accurate approximation? This threshold is difficult to determine. Bleistein (1966) devised a method which circumvents the need to choose a saddle-point. This method provides a complete asymptotic expansion of (2.12).

The method of Bleistein is applied in section 3.4 to derive an asymptotic expansion for the cumulative distribution function of a sample mean. The method can be generalised in order to extend its applicability to other cumulative distribution functions.

Bleistein's Method

Bleistein (1966) developed a method which yields an asymptotic estimate of the integral (2.12). His method hinges on the following three insights:

1. Perform the substitution $h(z) - h(0) = \frac{1}{2}w^2 - \hat{w}\,w$ (Bleistein 1966:362) which leads to the integral

$$\begin{aligned} I_\lambda &= \frac{1}{2\pi i}\int_a^b e^{\lambda\,h(z)}\,\frac{dz}{z^k} \\ &= \frac{e^{\lambda\,h(0)}}{2\pi i}\int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,\frac{dz}{dw}\,\frac{dw}{z^k} \\ &= \frac{e^{\lambda\,h(0)}}{2\pi i}\int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,G(w)\,\frac{dw}{w^k} \end{aligned} \tag{2.13}$$

where $G(w) = \frac{dz}{dw}\frac{w^k}{z^k}$ and a_1 and b_1 are the start- and end-points of the transformed path implied by the substitution (Bleistein, 1966:364). Here $\hat{w}$ is chosen as

$$\hat{w} = sign\{\hat{t}\}\,\sqrt{2}\,\sqrt{h(0) - h(\hat{t})} \tag{2.14}$$

where $sign\{\hat{t}\} = \frac{\hat{t}}{|\hat{t}|}$. The notation above differs slightly from that of Bleistein (1966:362) in the sense that $a = -\hat{w}$ and $\alpha = \hat{t}$.

2. Expand the term $G(w)$ as follows

$$G(w) = \frac{dz}{dw}\,\frac{w^k}{z^k} = c_0 + c_1\,w + w(w-\hat{w})\,G_1(w) \tag{2.15}$$

for some constants c_0 and c_1. By setting $w = 0$ and $w = \hat{w}$ it follows that the constants are given by

$$c_0 = G(0) = h(0)\left(\left.\frac{dz}{dw}\right|_{w=0}\right)^{1-k} \quad \text{and} \quad c_1 = \frac{G(\hat{w}) - c_0}{\hat{w}} = \frac{1}{\hat{w}}\left(h(\hat{w})\left(\frac{\hat{t}}{\hat{w}}\right)^{-k}\left.\frac{dz}{dw}\right|_{w=0} - c_0\right) \tag{2.16}$$

(Bleistein, 1966:364). The expansion (2.15) leads to the integral

$$\begin{aligned} I_\lambda &= \frac{e^{\lambda\,h(0)}}{2\pi i}\int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,G(w)\,\frac{dw}{w^k} \\ &= \frac{e^{\lambda\,h(0)}}{2\pi i}\left\{c_0\int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,\frac{dw}{w^k} + c_1\int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,\frac{dw}{w^{k-1}} + \int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,\epsilon(w)\,\frac{w-\hat{w}}{w^{k-1}}\,dw\right\} \\ &= e^{\lambda\,h(0)}\left\{c_0\,\lambda^{\frac{k-1}{2}}\,J_k\left(\sqrt{\lambda}\,\hat{w}\right) + c_1\,\lambda^{\frac{k-2}{2}}\,J_{k-1}\left(\sqrt{\lambda}\,\hat{w}\right) + J_{G_1}\right\} \end{aligned}$$

where

$$J_k(s) = \frac{1}{2\pi i} \int_{\sqrt{\lambda}\, a_1}^{\sqrt{\lambda}\, b_1} e^{\frac{1}{2}w^2 - sw} \frac{dw}{w^k} \tag{2.17}$$

$$J_{G_1} = \frac{1}{2\pi i} \int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2 - \hat{w}w\right)} G_1(w) \frac{w - \hat{w}}{w^{k-1}}\, dw$$

(Bleistein 1966:364). The integrals $J_k(s)$ and J_{G_1} are equal to $2\pi i\, U_r(s)$ and $2\pi i\, J_1(\lambda;\alpha)$ respectively where $U_r(s)$ and $J_1(\lambda;\alpha)$ are the integrals (6.15) and (6.18) of Bleistein(1966:364). The different forms of the integrals are as a result of the different forms of I_λ (given by equation (2.12)) and Bleistein's $I(\lambda;\alpha)$ (given by equation (6.1) (Bleistein,1966:362)).

3. The integral J_{G_1} is integrated by parts to obtain a recursive formula in terms of J_k and J_{k-1}. This recursive formula is applied to I_λ and yields the following asymptotic expansion:

$$I_\lambda = e^{\lambda h(0)} \left\{ \frac{J_k\left(\sqrt{\lambda}\,\hat{w}\right)}{\lambda^{\frac{-k+1}{2}}} \left(\sum_{j=0}^{n} \frac{c_{2j}}{\lambda^j} + O\left(\lambda^{-n-1}\right) \right) + \frac{J_{k-1}\left(\sqrt{\lambda}\,\hat{w}\right)}{\lambda^{\frac{-k+2}{2}}} \left(\sum_{j=0}^{n} \frac{c_{2j+1}}{\lambda^j} + O\left(\lambda^{-n-1}\right) \right) \right\} \tag{2.18}$$

as $\lambda \to \infty$ for any $n = 0, 1, 2, \ldots$ and some constants $c_0, c_1, c_2, \ldots$ (Bleistein, 1966:365). The constants c_0 and c_1 are given by equation (2.16) and the remaining constants $c_2, c_3, \ldots$ are given by recursive equation (7.3) from Bleistein (1966:365).

The three insights of Bleistein can be summarised as follows

$$\begin{aligned}
I_\lambda &= \frac{1}{2\pi i} \int_a^b e^{\lambda h(z)} \frac{dz}{z^k} \\
&= \frac{e^{\lambda h(0)}}{2\pi i} \int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2 - \hat{w}w\right)} G(w) \frac{dw}{w^k} \\
&= e^{\lambda h(0)} \left\{ c_0\, \lambda^{\frac{k-1}{2}}\, J_k\left(\sqrt{\lambda}\,\hat{w}\right) + c_1\, \lambda^{\frac{k-2}{2}}\, J_{k-1}\left(\sqrt{\lambda}\,\hat{w}\right) + J_{G_1} \right\} \\
&= e^{\lambda h(0)} \left\{ \frac{J_k\left(\sqrt{\lambda}\,\hat{w}\right)}{\lambda^{\frac{-k+1}{2}}} \left(\sum_{j=0}^{n} \frac{c_{2j}}{\lambda^j} + O\left(\lambda^{-n-1}\right) \right) + \frac{J_{k-1}\left(\sqrt{\lambda}\,\hat{w}\right)}{\lambda^{\frac{-k+2}{2}}} \left(\sum_{j=0}^{n} \frac{c_{2j+1}}{\lambda^j} + O\left(\lambda^{-n-1}\right) \right) \right\}
\end{aligned}$$

as $\lambda \to \infty$ for any $n = 0, 1, 2, \ldots$ and some constants $c_0, c_1, c_2, \ldots$ where c_0 and c_1 are given by equation (2.16) and $J_k(s)$ is given by equation (2.17).

The Fourier inversion formula of the cumulative distribution function of a sample mean can be written in the form of (2.12) with $k = 1$. If $k = 1$ then the singularity at the origin is called a simple pole. This case is investigated in the following section and the corresponding results are used in section 3.4.6 to derive the saddle-point cumulative distribution function of a sample mean.

The Case of a Simple Pole at the Origin

By choosing $k = 1$ it follows from equation (2.18) that Bleistein's method yields

$$\int_a^b e^{\lambda h(z)} \frac{dz}{z} = e^{\lambda h(0)} \left\{ J_k\left(\sqrt{\lambda}\,\hat{w}\right) \left(\sum_{j=0}^{n} \frac{c_{2j}}{\lambda^j} + O\left(\lambda^{-n-1}\right) \right) + J_{k-1}\left(\sqrt{\lambda}\,\hat{w}\right) \left(\sum_{j=0}^{n} \frac{c_{2j+1}}{\lambda^{j+\frac{1}{2}}} + O\left(\lambda^{-n-\frac{3}{2}}\right) \right) \right\}$$
$$= e^{\lambda h(0)} \left\{ J_1\left(\sqrt{\lambda}\,\hat{w}\right) \left(c_0 + O\left(\lambda^{-1}\right)\right) + \frac{1}{\sqrt{\lambda}} J_0\left(\sqrt{\lambda}\,\hat{w}\right) \left(c_1 + O\left(\lambda^{-1}\right)\right) \right\} \qquad (2.19)$$

where

$$J_0(s) = \frac{1}{2\pi i} \int_{\sqrt{\lambda}\,a_1}^{\sqrt{\lambda}\,b_1} e^{\frac{1}{2}w^2 - ws}\, dw \qquad \text{and} \qquad J_1(s) = \frac{1}{2\pi i} \int_{\sqrt{\lambda}\,a_1}^{\sqrt{\lambda}\,b_1} e^{\frac{1}{2}w^2 - ws}\, \frac{dw}{w}$$

$$c_0 = 1 \qquad \text{and} \qquad c_0 = \frac{1}{\hat{t}\,\sqrt{h^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}$$

The Fourier inversion formula of the cumulative distribution function of a sample mean can be written in the form of (2.12) with $k = 1$. Result (2.19) can therefore be applied to yield a complete asymptotic expansion for the cumulative distribution function of a sample mean.

The cumulative distribution functions of other statistics, however, are often not of the form of (2.12). It is therefore essential to derive a saddle-point method for a more general integral. The saddle-point method for this general integral, in turn, has wider applications than does (2.19).

A More General Integral

Consider the integral

$$I_\lambda = \frac{1}{2\pi i} \int_a^b g_\lambda(z) \frac{dz}{z} = \frac{1}{2\pi i} \int_a^b e^{h_\lambda(z)} \frac{dz}{z} \qquad (2.20)$$

where $g_\lambda(z) = e^{h_\lambda(z)}$ is analytic on an open set $\Omega \subset \mathbb{C}$, the path of integration is a piece-wise smooth path from a to b in Ω and the parameter λ is large.

The substitution $h_\lambda(z) - h_\lambda(0) = \frac{1}{2}w^2 - \hat{w}\,w$ is performed which yields the integral

$$I_\lambda = \frac{e^{\lambda h(0)}}{2\pi i} \int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2 - \hat{w}w\right)} G(w) \frac{dw}{w}$$

where $G(w) = \frac{dz}{dw}\frac{w^k}{z^k}$ and a_1 and b_1 are the start- and end-points of the transformed path implied by the substitution. Here $\hat{w}$ is chosen as

$$\hat{w} = sign\{\hat{t}\}\,\sqrt{2}\,\sqrt{h(0) - h(\hat{t})}$$

It is required that $G(w)$ can be expanded as follows for all w on the path from a_1 to b_1.:

$$G(w) \sim G(0) + \frac{G(\hat{w}) - G(0)}{\hat{w}}\, w = 1 + \left\{ \frac{1}{\hat{t}\,\sqrt{h^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right\} w \qquad \text{as } \lambda \to \infty \qquad (2.21)$$

The condition on $G(w)$ is inspired by the expansion (2.15) (where $k = 1$) which leads to the result (2.19). If the condition on $G(w)$ is satisfied then

$$\begin{aligned} I_\lambda &= \frac{1}{2\pi i}\int_a^b e^{h_\lambda(z)}\,dz \\ &\sim \frac{e^{\lambda\,h(0)}}{2\pi i}\int_{a_1}^{b_1} e^{\lambda\left(\frac{1}{2}w^2-\hat{w}w\right)}\,G(w)\,\frac{dw}{w} \qquad \text{by the substitution} \\ &\sim e^{h_\lambda(0)}\,\{c_0\,J_1(\hat{w})+c_1\,J_0(\hat{w})\} \qquad \text{by the condition on } G(w) \\ &= e^{\lambda\,h(0)}\left\{J_1\left(\hat{w}\right)+\left(\frac{1}{\hat{t}\,\sqrt{h^{(2)}(\hat{t})}}-\frac{1}{\hat{w}}\right)\,J_0\left(\hat{w}\right)\right\} \qquad \text{as } \lambda\to\infty \end{aligned} \tag{2.22}$$

where

$$J_0(s)=\frac{1}{2\pi i}\int_{\sqrt{\lambda}\,a_1}^{\sqrt{\lambda}\,b_1} e^{\frac{1}{2}w^2-ws}\,dw \qquad \text{and} \qquad J_1(s)=\frac{1}{2\pi i}\int_{\sqrt{\lambda}\,a_1}^{\sqrt{\lambda}\,b_1} e^{\frac{1}{2}w^2-ws}\,\frac{dw}{w} \tag{2.23}$$

Result (2.22) is similar to result (2.19) but there are several important differences. Result (2.22) has wider applications than does (2.19), but does not provide a complete asymptotic expansion of I_λ nor does it provide the order by which I_λ converges. It is necessary to know $h_\lambda(z)$ in order to determine the asymptotic expansion of I_λ (if one exists). Result (2.22) does, however, provide the asymptotic limit of I_λ as $\lambda\to\infty$ given that the condition (2.21) on $G(w)$ is satisfied.

2.4 Summary

2.4.1 Introduction

This section aims to capture the key points discussed in this chapter. The discussion of these key points are categorised according to the section in which they first appear, since many of the points feature in multiple sections.

2.4.2 Complex Analysis

The saddle-point method is a complex analysis based approach to deriving an asymptotic estimate of a complex integral, and a complex integral depends on two items explicitly: the path of integration ϕ and the function being integrated g. The aim of section 2.2 is to provide background to complex analysis, specifically concerning paths and complex functions. These two subjects are briefly summarised below.

A path ϕ is a continuous complex function of a real variable and is written as $\phi : [s_1, s_2] \to \mathbb{C}$ where $[s_1, s_2] \subset \mathbb{R}$. A complex integral is evaluated along a path. One of important properties of complex integration is that a path may be deformed as long as it has the same start- and end-points and does not cross a singularity.

A complex function of a complex variable $g(z)$ is of the form $g(x + iy) = u(x, y) + i\,v(x, y)$ where $x, y \in \mathbb{R}$ and $u, v : \mathbb{R}^2 \to \mathbb{R}$. The function g is said to be analytic on an open set $\Omega \subset \mathbb{C}$ if it is infinitely differentiable on Ω. Analytic functions have points of only three possible types: zero points, ordinary points and saddle-points. The saddle-point method assumes that the function being integrated is analytic over the path of integration.

The singularities of the function g are the set of complex points where the function is not analytic. Singularities play a central role in complex integration and in the saddle-point method. Singularities create saddle-points. By referring back to the analogy of the mountain range, singularities form the unending mountain peaks which, in turn, form the mountain passes which are synonymous to saddle-points.

2.4.3 The Saddle-point Method

The Saddle-point Methods of de Bruijn (1958) and Bleistein (1966)

Two saddle-point methods are discussed in section 2.3, namely the saddle-point method of de Bruijn (1958) and the saddle-point method of Bleistein (1966). These two methods can be applied to the Fourier inversion formula of the density and distribution function of the sample mean respectively to yield the saddle-point density and distribution functions of said sample mean. This is discussed in the following chapter in sections 3.2.7 and 3.3.6. The saddle-point methods of de Bruijn (1958) and Bleistein (1966) are briefly summarised below.

The saddle-point method of de Bruijn (1958:87) has two parts: the choice of path and Laplace's method. This method is derived by deforming the path of integration into a steepest descent path and performing a substitution which makes the integral bounds real. Laplace's method is applied to this integral and produces the simple saddle-point asymptotic expansion of equation (2.9).

This saddle-point method of Bleistein (1966) is developed for a complex integral whose function has a singularity at the origin. This method entails performing a substitution and expanding the term $G(w)$ linearly so that the original integral becomes the asymptotic expansion given by equation (2.18).

More Generalised Saddle-point Methods

The saddle-point methods of de Bruijn (1958) and Bleistein (1966) are generalised in order to extend their applicability.

The saddle-point method of de Bruijn (1958) is generalised by Flajolet & Sedgewick (2009). In the saddle-point method of Flajolet & Sedgwick (2009) it is required that the function being integrated depends on a large parameter λ such that three conditions are satisfied. Under these conditions the complex integral (2.10) converges to the function given in equation (2.11). By adapting one of the conditions of Flajolet & Sedgewick (2009), their method can be applied to the Fourier inversion formula of any density function with a well-behaved characteristic function. This is discussed in the following chapter in section 3.3.

The method of Bleistein is generalised by requiring that the function being integrated depends on a large parameter λ such that the function satisfies the condition (2.21). Under this condition the complex integral converges to the function given in equation (2.22). By adapting this condition, the generalised method of Bleistein can be applied to the Fourier inversion formula of any distribution function with a well-behaved characteristic function. This is discussed in the following chapter in section 3.4.

Chapter 3

The Saddle-point Approximation

3.1 Introduction

The application of the saddle-point method to the Fourier inversion formula yields the saddle-point approximation. The approximation is introduced into statistics by Daniels (1954) who applies a saddle-point method similar to that of de Bruijn (1958) to the Fourier inversion formula of the density function of a sample mean. This yields a uniform asymptotic expansion of the density function for the sample mean, which is aptly named "the saddle-point approximation". Lugannani & Rice (1980) use a saddle-point method similar to that of Bleistein (1966) on the Fourier inversion formula of the distribution function of a sample mean. This yields a uniform asymptotic expansion for the distribution function of the sample mean, which is called the formula of Lugannani & Rice. Both of these results are asymptotic. However, due to their low relative error rate, the saddle-point approximation and the formula of Lugannani & Rice are often used in a non-asymptotic setting. The saddle-point approximation of Daniels (1954) and the formula of Lugannani & Rice (1980) are discussed in sections 3.2.7 and 3.3.6 of this chapter.

Hougaard (1988) is the first to use the saddle-point approximation to approximate the density function of a random variable which does not depend on sample size, or any other large parameter. This opens up the possibility of using the saddle-point approximation in cases where no large parameters are involved, i.e. a non-asymptotic setting. Due to its high accuracy in the non-asymptotic setting, the saddle-point approximation currently refers to the general formulas of the saddle-point density and distribution functions as given by Butler (2007). The theory underlying these general formulas, however, is often lacking or non-existent. The primary aim of this chapter is to discuss and propose possible adaptations of the saddle-point methods of chapter 2 in order for them to yield the saddle-point density and distribution functions of Butler (2007).

This chapter has three parts. The first part comprises of a discussion of the saddle-point density function. The second part entails a discussion of the saddle-point distribution function. The final part serves as a summary of the chapter.

3.2 The Saddle-point Approximation of a Density Function

3.2.1 Introduction

In this section, the saddle-point method of section 2.3.5 is adapted and applied to the Fourier inversion formula of a probability density function. This yields an approximation of the given probability density function, which is aptly called the saddle-point density function.

The saddle-point method of section 2.3.5 is applicable to an integral which has a large parameter λ. Many probability density functions, however, have no large parameters. In the case of such a density function, the saddle-point method still yields an accurate approximation. The accuracy of the saddle-point density function in cases where there is no large parameter is counter-intuitive - an asymptotic method yields accurate results in non-asymptotic cases. This phenomenon is one of the central topics in this thesis and is expanded on in this section.

This section has four parts. Firstly, the Fourier inversion formula is discussed. Secondly, the saddle-point method of section 2.3.5 is adapted and applied to approximate the density function of a random variable with no large parameter. The resulting approximation is called the saddle-point density function. Thirdly, an example is given of the saddle-point density function of a non-central chi-squared random variable. Finally, the saddle-point asymptotic expansion of the density function a sample mean is derived using the saddle-point method of de Bruijn (1958). This asymptotic expansion is originally derived by Daniels (1954).

3.2.2 The Fourier Inversion Formula

Consider a random variable X with distribution function $F(x) = P(X \leq x)$ and cumulant generating function $K(t) = \log E[e^{tX}]$. It is a well-known result (Lukacs, 1970:33) that the density function $f(x) = \frac{dF(x)}{dx}$ is given by the Fourier inversion formula:

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{K(it)-itx}\,dt \\ &= \frac{1}{2\pi}\lim_{r\to\infty}\int_{-ir}^{ir} e^{K(z)-zx}\,dz \\ &= \frac{1}{2\pi i}\lim_{r\to\infty}\int_{\phi_r} e^{K(z)-zx}\,dz \end{aligned} \tag{3.1}$$

where $\phi_r(s) = is$ for $-r \leq s \leq r$ is the straight path from $-ir$ to ir. This integral has the same form as the integral in equation (2.10) where $h(z) = K(z) - zx$ but with no large parameter λ present.

3.2.3 The Saddle-point Density Function

The basic saddle-point method of Flajolet & Sedgewick (2009), which is discussed in section 2.3.5, is applicable to an integral (2.10) where $h_\lambda(z)$ depends on a large parameter λ. Although the Fourier inversion formula given by (3.1) is of the same form as (2.10), the function $h(z) = K(z) - zx$ does not depend on a large parameter. This implies that the saddle-point method of Flajolet & Sedgewick (2009) cannot be applied directly to equation (3.1) because the second condition of Flajolet & Sedgewick (2009) is not satisfied. The first and third conditions

of Flajolet & Sedgewick (2009) are satisfied automatically by choosing the path of steepest descent as a straight path. This is discussed later in this section.

In order to apply an adapted version of the basic saddle-point method to (3.1), it is necessary to adjust the second condition of Flajolet & Sedgewick (2009) for the case where there is no large parameter. This adapted saddle-point method is then applied to equation (3.1) to yield the saddle-point density function. The saddle-point density function is simply an approximation and not an asymptotic result if there is no large parameter present.

The derivation of the saddle-point density function is discussed in four parts. Firstly, the saddle-point is defined and its properties are discussed. Secondly, the path of steepest descent is chosen. Thirdly, the three conditions of Flajolet & Sedgewick (2009) are discussed. Finally, the saddle-point density function is obtained.

The Saddle-point

The saddle-point $\hat{t}$ of $e^{K(z)-zx}$ is defined in section 2.2.5 as the solution to the equation

$$K^{(1)}(\hat{t}) = x \qquad \text{such that} \quad \hat{t} \in (t_1, t_2)$$

where $(t_1, t_2) \subset \mathbb{R}$ is the interval on which $K(t)$ is defined. Daniels (1954:637-639) proves that $\hat{t}$ exists, that it is unique and that $K^{(2)}(\hat{t}) > 0$.

The Path of Steepest Descent

The saddle-point $\hat{t}$ is real and so is $K^{(2)}(\hat{t})$. The argument of the saddle-point axis is therefore $\frac{\pi}{2}$ which follows from its definition in section 2.3.4. This implies that the saddle-point axis is the straight vertical line that intercepts the real axis at the saddle-point $\hat{t}$.

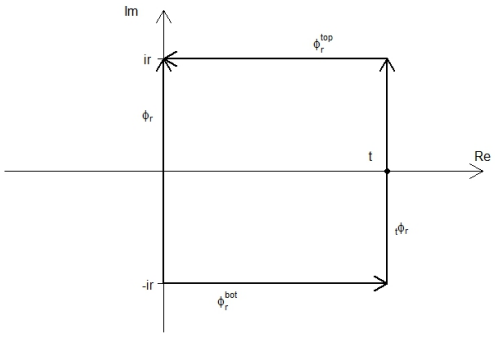
A path of steepest descent crosses the real axis at the saddle-point perpendicularly. An optimal path of steepest descent is given by $\hat{\phi}_r(s) = \hat{t} + is$ for $-r \leq s \leq r$. The reason for this choice is elaborated upon under the following heading.

The Conditions

Application of the saddle-point method entails choosing the path of integration as a path of steepest descent and adhering to the three conditions of Flajolet & Sedgewick (2009) from section 2.3.5.

The path ϕ_r of the integral (3.1) can be deformed into the combined path $\phi_r^{bot} \cup \hat{\phi}_r \cup \phi_r^{top}$ where ϕ_r^{bot} is the straight path from $-ir$ to $-ir + \hat{t}$, $\hat{\phi}_r$ is the straight path from $-ir + \hat{t}$ to $ir + \hat{t}$ and ϕ_r^{top} is the straight path from $ir + \hat{t}$ to ir. The steepest descent path $\hat{\phi}_r$ is optimal since it eliminates condition 3 of Flajolet & Sedgewick (2009). This is demonstrated later in this section in equation (3.3).

The cumulant generating function is real and therefore has real singularites (if there are any) which lie outside its domain (t_1, t_2). The landscape $Re\{K(z) - zx\}$ forms a symmetric mountain range along the real axis with all the unending peaks lying on the real axis.



It follows from Daniels (1954:632) that the sum of the integrals over the tail paths ϕ_r^{bot} and ϕ_r^{top} equals 0 as $r \to \infty$. Therefore condition 1 of Flajolet & Sedgewick (2009) is automatically satisfied. There are two intuitive reasons why the tail paths become negligible for large r. Firstly, the landscape $Re\{K(z) - zx\}$ becomes flatter further away from the real axis. The tail paths ϕ_r^{bot} and ϕ_r^{top} run parallel on either side of the real axis along this flat landscape. Therefore, the larger r becomes, the further away the tail paths are from the real axis and the flatter the landscape along the tail paths become. Secondly, the tail paths run in opposite parallel directions on either side of the real axis and the landscape $Re\{K(z) - zx\}$ is symmetric around the real axis. This implies that the integrals along the two tail paths offset each other.

The final condition to be satisfied is an adjustment of condition 2 of Flajolet & Sedgewick (2009): that the quadratic approximation

$$K(\hat{t} + ir) - x\left(\hat{t} + ir\right) \approx K(\hat{t}) - \hat{t}\,x - \frac{1}{2}\,K^{(2)}(\hat{t})\,r^2 \tag{3.2}$$

is uniformly adequate for $-\infty < r < \infty$.

This approximation is adequate for distributions with well-behaved cumulant generating functions.

The Saddle-point Density Function

If condition (3.2) is satisfied, it follows that

$$\begin{aligned}
f(x) &= \frac{1}{2\pi i} \lim_{r\to\infty} \int_{\phi_r} e^{K(z)-zx}\,dz \\
&= \frac{1}{2\pi i} \lim_{r\to\infty} \int_{\hat{\phi}_r} e^{K(z)-zx}\,dz + \frac{1}{2\pi i} \lim_{r\to\infty} \left(\int_{\phi_r^{bot}} e^{K(z)-zx}\,dz + \int_{\phi_r^{top}} e^{K(z)-zx}\,dz \right) \\
&= \frac{1}{2\pi i} \lim_{r\to\infty} \int_{\hat{\phi}_r} e^{K(z)-zx}\,dz \\
&\approx \frac{1}{2\pi i}\, e^{K(\hat{t})-\hat{t}\,x} \lim_{r\to\infty} \int_{\hat{\phi}_r} e^{\frac{1}{2}\,K^{(2)}(\hat{t})\,(z-\hat{t})^2}\,dz \\
&= \frac{1}{2\pi}\, e^{K(\hat{t})-\hat{t}\,x} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\,K^{(2)}(\hat{t})\,s^2}\,ds \\
&= \frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{2\pi\,K^{(2)}(\hat{t})}}
\end{aligned} \tag{3.3}$$

which is called the saddle-point density function (Butler, 2007:3).

3.2.4 Remarks

No Large Parameter

The saddle-point density function is an approximation and can be determined for any distribution which has a well-behaved cumulant generating function $K(t)$. The accuracy of the saddle-point density function depends on the adequacy of quadratic approximation

$$K(\hat{t}+ir)-x\left(\hat{t}+ir\right)\approx K(\hat{t})-\hat{t}\,x-\frac{1}{2}K^{(2)}(\hat{t})\,r^2 \qquad \text{for} \qquad r\in\mathbb{R}$$

The saddle-point density serves as a useful (closed-form) approximation of a non-closed-form probability density function. An example is the non-central chi-squared density function which is discussed in section 3.2.6.

A Large Parameter

The saddle-point density is asymptotic if the cumulant generating function depends on some large value of λ such that

$$K(\hat{t}+ir)-x\left(\hat{t}+ir\right)=K(\hat{t})-\hat{t}\,x-\frac{1}{2}K^{(2)}(\hat{t})\,r^2+O(\epsilon_\lambda)$$

where $\epsilon_\lambda\to 0$ as $\lambda\to\infty$ uniformly for $r\in\mathbb{R}$. Flajolet & Sedgewick (2009: 552) prove that under these circumstances

$$f(x)\sim\frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{2\pi\,K^{(2)}(\hat{t})}} \qquad \text{as } \lambda\to\infty$$

Furthermore, the saddle-point method yields a complete asymptotic expansion for the density function if the Fourier inversion formula is of the form of equation (2.5). The sample mean is an example of such a case and is further discussed in section 3.2.7.

3.2.5 Renormalising the Saddle-point Density Function

The normalised saddle-point density function $\hat{f}(x)$ corresponding to the cumulant generating function $K(t)$ is defined as

$$\hat{f}(x)=\frac{1}{c}\frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{K^{(2)}(\hat{t})}} \qquad \text{where} \qquad c=\int_{-\infty}^{\infty}\frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{K^{(2)}(\hat{t})}}\,dx=\int_{t_1}^{t_2}\frac{e^{K(t)-t\,K^{(1)}(t)}}{\sqrt{K^{(2)}(t)}}\,dt$$

given that $K(t)$ is defined on the domain $(t_1,t_2)\subset\mathbb{R}$. The substitution $K^{(1)}(t)=x$ in the integral c is recommended by Butler (2007:5) and generally leads to an easier integral since t_1 and/or t_2 is finite. The renormalised saddle-point density function is a probability density function.

3.2.6 Example: The Non-central Chi-squared Density Function

Hougaard (1988) demonstrates that the saddle-point density serves as an accurate approximation of the non-central chi-squared density function. The non-central chi-squared distribution, with degrees of freedom p and non-centrality parameter η, has the moment generating function $M(t)=e^{\frac{\eta t}{1-2t}}(1-2t)^{-\frac{p}{2}}$. It follows that $K(t)=\frac{\eta t}{1-2t}-\frac{p}{2}\log(1-2t)$ and $K^{(1)}(t)=\frac{\eta}{(1-2t)^2}+\frac{p}{1-2t}$.

The saddle-point $\hat{t}$ is given by

$$\begin{aligned} K^{(1)}(\hat{t}) &= x \\ \eta + p(1-2\hat{t}) &= (1-4\hat{t}+4\hat{t}^2)x \\ \eta + p - x &= 4x\hat{t}^2 + (2p-4x)\hat{t} \\ \hat{t} &= \frac{2x - p \pm \sqrt{p^2+4\eta x}}{4x} \end{aligned}$$

There are two saddle-points, say $\hat{t}_+$ and $\hat{t}_-$. One of the conditions is that the saddle-point should lie in the domain of $K(t)$. $K(t)$ has a singularity at $t = \frac{1}{2}$. If $\eta = 0$ then $\hat{t}_+ = \frac{2x-p+\sqrt{p^2+4\eta x}}{4x} = \frac{1}{2}$, which does not satisfy the condition. Therefore we will only consider the saddle-point $\hat{t}(x) = \hat{t}_- = \frac{2x-p-\sqrt{p^2+4\eta x}}{4x}$ which does satisfy the condition for any $p > 0$ and $\eta \geq 0$.

To simplify matters, let $r(x) = \frac{p+\sqrt{p^2+4\eta x}}{2x}$ then $\hat{t}(x) = \frac{1-r(x)}{2}$ and it follows that

$$\begin{aligned} K(\hat{t}(x)) - \hat{t}(x)x &= \frac{1-r(x)}{2}\left(\frac{\eta}{r(x)} - x\right) - \frac{p}{2}\log r(x) \\ K^{(2)}(\hat{t}(x)) &= \frac{4\eta}{r^3(x)} + \frac{2p}{r^2(x)} = \frac{2}{r^2(x)}\left(\frac{2\eta}{r(x)} + p\right) \end{aligned}$$

The non-central chi-squared saddle-point density is therefore given by

$$f(x) \approx \frac{e^{K\left(\hat{t}(x)\right) - \hat{t}(x)x}}{\sqrt{2\pi K^{(2)}\left(\hat{t}(x)\right)}} = \frac{r^{1-\frac{p}{2}}(x) e^{\frac{1-r(x)}{2}\left(\frac{\eta}{r(x)} - x\right)}}{2\sqrt{\pi}\sqrt{\frac{2\eta}{r(x)} + p}}$$

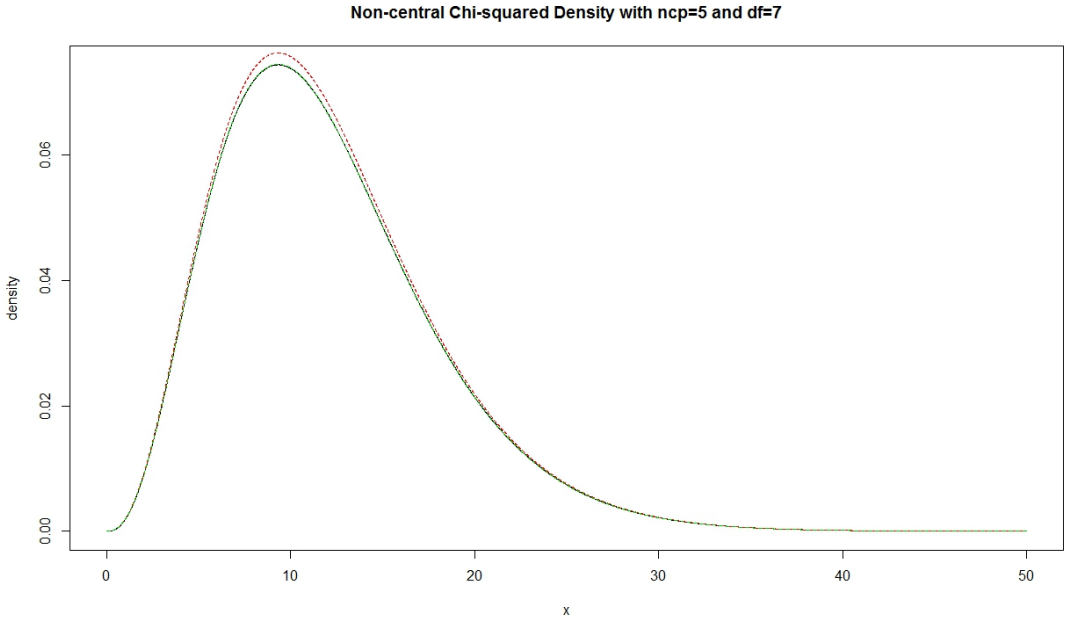


Figure 3.1: The exact density is given by the solid black line, the saddle-point density function is given by the dashed red line and the normalised saddle-point density is given by the dashed green line. The normalised saddle-point density lies on the exact density and therefore provides an almost-perfect fit.

3.2.7 The Saddle-point Density Function of a Sample Mean

Consider a random sample $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables with the distribution function $F(x) = P(X \leq x)$ and cumulant generating function $K(t) = \log E\left[e^{tX}\right]$. Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ as the sample mean. The density function of the sample mean, say $f_n(\bar{x})$, is given by the Fourier inversion formula:

$$\begin{aligned} f_n(\bar{x}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{n\,K\left(\frac{it}{n}\right)-it\bar{x}}\,dt \\ &= \frac{n}{2\pi}\int_{-\infty}^{\infty} e^{n(K(is)-is\bar{x})}\,ds \\ &= \frac{n}{2\pi i}\lim_{r\to\infty}\int_{\phi_r} e^{n(K(z)-z\bar{x})}\,dz \end{aligned} \tag{3.4}$$

which is of the form (2.5) where $h(z) = K(z) - z\,x$ and $\lambda = n$ is large. By applying (2.9) it follows that

$$\begin{aligned} f_n(\bar{x}) &= \frac{n}{2\pi i}\lim_{r\to\infty}\int_{\hat{\phi}_r} e^{n(K(z)-z\bar{x})}\,dz \\ &= \frac{\sqrt{n}\,e^{n\left(K(\hat{t})-\hat{t}\,\bar{x}\right)}}{\sqrt{2\pi\,K^{(2)}(\hat{t})}}\left(1+\frac{c_1}{n}+\frac{c_2}{n^2}+\ldots\right) \\ &= \frac{\sqrt{n}\,e^{n\left(K(\hat{t})-\hat{t}\,\bar{x}\right)}}{\sqrt{2\pi\,K^{(2)}(\hat{t})}}\left(1+O\left(n^{-1}\right)\right) \qquad \text{as } \lambda\to\infty \end{aligned} \tag{3.5}$$

where $c_1, c_2, \ldots$ are real constants that depend on $\hat{t}$.

The asymptotic expansion for the density function of a sample mean (given by equation (3.5)) can be derived using the Edgeworth expansion indirectly. This derivation is devised by Daniels (1954) who introduced the saddle-point method into statistics.

3.3 The Saddle-point Approximation of a Distribution Function

3.3.1 Introduction

In this section, the saddle-point method of section 2.3.6 is applied to the Fourier inversion formula of a cumulative distribution function. This yields an approximation of the given cumulative distribution function, which is aptly called the saddle-point distribution function.

The saddle-point method of section 2.3.6 is applicable to an integral which has a large parameter λ. Many probability distribution functions, however, have no large parameters. In the case of such a distribution function, the saddle-point method still yields an accurate approximation. This phenomenon, which is mentioned in section 3.2, is one of the central topics in this thesis and is further discussed in this section.

This section has four parts. Firstly, the inverse Fourier transformation is discussed. Secondly, the saddle-point method of section 2.3.6 is applied to approximate the cumulative distribution function of a random variable with no large parameter. The resulting approximation is called the saddle-point distribution function. Thirdly, an example is given of the saddle-point distribution function of a non-central chi-squared random variable. Finally, the saddle-point asymptotic expansion for the distribution function a sample mean is derived using Bleistein's saddle-point method. This asymptotic expansion is originally derived by Lugannani & Rice (1980).

3.3.2 The Fourier Inversion Formula

Consider a random variable X with distribution function $F(x) = P(X \leq x)$ and cumulant generating function $K(t) = \log E[e^{tX}]$. It follows from Daniels (1987:39) that the survival function $1 - F(x)$ can be written as the Fourier inversion formula:

$$1 - F(x) = \frac{1}{2\pi i} \lim_{r\to\infty} \int_{t\phi_r} e^{K(z)-zx} \frac{dz}{z} \tag{3.6}$$

where ${}_t\phi_r(s) = t + is\,,\ -r \leq s \leq r$ is the straight path from $t - ir$ to $t + ir$ and $t > 0$ lies in the domain of $K(t)$.

3.3.3 The Saddle-point Distribution Function

The saddle-point method which compensates for a simple pole at the origin (discussed in section 2.3.6) is applicable to an integral (3.6) where $h_\lambda(z)$ depends on a large parameter λ. Although the Fourier inversion formula given by (3.6) is of the same form as (2.20), the function $h(z) = K(z) - zx$ does not depend on a large parameter. This implies that the saddle-point method (resulting in equation (2.22)) cannot be applied directly to equation (3.6) because the condition (2.21) on $G(w)$ is not satisfied.

In order to apply an adapted version of the saddle-point method which compensates for a simple pole at the origin, it is necessary to adjust the condition on $G(w)$ for the case where there is no large parameter. The adapted saddle-point method is applied to equation (3.6) to yield the saddle-point distribution function. The saddle-point distribution function is simply

an approximation and not an asymptotic result if there is no large parameter present.

The derivation of the saddle-point distribution function is discussed in three parts. Firstly, the adjusted condition on $G(w)$ is discussed. Secondly, the integrals $J_1(s)$ and $J_0(s)$ are determined. Finally, the saddle-point distribution function is obtained.

The Condition on $G(w)$

Consider the transformation $\frac{1}{2}w^2 - \hat{w}\,w = K(z) - z\,x$ where $\hat{w}$ is chosen as

$$\hat{w} = sign\{\hat{t}\}\,\sqrt{2}\,\sqrt{\hat{t}\,x - K(\hat{t})}$$

so that w and z behave similarly around $w = 0$ and $w = \hat{w}$ (Daniels, 1987:42).

The condition on $G(w)$, which is an adaptation of (2.21), is given as follows:

$$G(w) = \frac{dz}{dw}\,\frac{w}{z} \approx G(0) + \frac{G(\hat{w}) - G(0)}{\hat{w}}\,w = 1 + \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right) w \tag{3.7}$$

is uniformly adequate for all w on the path ${}_t\phi_r$.

The fact that $G(0) = 1$ and $G(\hat{w}) = \frac{\hat{w}}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}}$ follows from the choice of $\hat{w}$.

Condition (3.7) can be seen as a condition on the transformation $\frac{1}{2}w^2 - \hat{w}\,w = K(z) - z\,x$ so that $G(w) = \frac{dz}{dw}\,\frac{w}{z}$ is approximately linear in w so that $G(0) = 1$ and $G(\hat{w}) = \frac{1}{\sqrt{K^{(2)}(\hat{t})}}\,\frac{\hat{w}}{\hat{t}}$. This condition is, however, much less intuitive than condition (3.2).

If condition (3.7) is satisfied then it follows that

$$\begin{aligned}
1 - F(x) &= \frac{1}{2\pi i}\lim_{r\to\infty}\int_{{}_t\phi_r} e^{K(z)-zx}\,\frac{dz}{z} \\
&= \frac{1}{2\pi i}\lim_{r\to\infty}\int_{{}_t\phi_r} e^{\frac{1}{2}w^2-\hat{w}\,w}\,\frac{dz}{dw}\frac{dw}{z} \\
&\approx \frac{1}{2\pi i}\lim_{r\to\infty}\int_{{}_t\phi_r} e^{\frac{1}{2}w^2-\hat{w}\,w}\,\frac{dw}{w} + \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right)\frac{1}{2\pi i}\lim_{r\to\infty}\int_{{}_t\phi_r} e^{\frac{1}{2}w^2-\hat{w}\,w}\,dw \\
&= J_1(\hat{w}) + J_0(\hat{w})\left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right)
\end{aligned} \tag{3.8}$$

where the integrals $J_1(s)$ and $J_0(s)$ are defined by equation (2.23) in section 2.3.6. It follows from Daniels (1987:42) that the transformation $K(z) - zx = \frac{1}{2}w^2 - \hat{w}w$ does not change the path of integration ${}_t\phi_r$. This is because the transformation from z to w is surjective, monotone-increasing and continuous (Butler, 2007:51). Therefore both the integrals $J_1(s)$ and $J_2(s)$ are evaluated over the path ${}_t\phi_r$.

The Integrals $J_1(\hat{w})$ and $J_0(\hat{w})$

The cumulant generating function of the standard normal distribution is given by $K_\Phi(w) = \frac{1}{2}\,w^2$. From equations (3.1) and (3.6) it follows that the integrals $J_1(\hat{w})$ and $J_0(\hat{w})$ along the

path ${}_t\phi_r$ are the standard normal survival function and density function respectively. In other words,

$$J_1(\hat{w}) = \int_{{}_t\phi_r} e^{\frac{1}{2}w^2 - \hat{w}\,w}\,\frac{dw}{w} = 1 - \Phi(\hat{w})$$
$$J_0(\hat{w}) = \int_{{}_t\phi_r} e^{\frac{1}{2}w^2 - \hat{w}}\,dw = \phi(\hat{w})$$

By substituting this result into equation (3.8) it follows that

$$1 - F(x) \approx J_1(\hat{w}) - J_0(\hat{w})\left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right) = \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right)$$

The Saddle-point Distribution Function

If condition (3.7) is satisfied then the above result can be rewritten as

$$F(x) \approx \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right) \tag{3.9}$$

It is important to note that $\hat{t} = \hat{w} = 0$ when $x = E[X]$. However, the term $\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}$ is continuous in the region around $x = E[X]$ (Butler, 2007:12). It follows from Butler (2007:69) that

$$\lim_{\hat{t}\to 0}\left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right) = \frac{1}{6}\frac{K^{(3)}(0)}{(K^{(2)}(0))^{\frac{3}{2}}}$$

which implies that equation (3.9) can be written as

$$F(x) \approx \begin{cases} \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right) & \text{if } x \neq E[X] \\ \frac{1}{2} + \frac{1}{6\sqrt{2\pi}}\frac{K^{(3)}(0)}{\left(K^{(2)}(0)\right)^{\frac{3}{2}}} & \text{if } x = E[X] \end{cases} \tag{3.10}$$

which is called the saddle-point distribution function (Butler, 2007:12).

3.3.4 Remarks

No Large Parameter

The saddle-point distribution function is an approximation and can be determined for any distribution which has a well-behaved cumulant generating function $K(t)$. The accuracy of the saddle-point distribution function depends on the adequacy of the linear approximation

$$\frac{dz}{dw}\,\frac{w}{z} \approx 1 + \left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}\right)w \qquad \text{for} \qquad z \in {}_t\phi_r^*$$

The saddle-point distribution function serves as a useful (closed-form) approximation of a non-closed-form probability distribution function. An example of is the non-central chi-squared distribution function which is discussed in section 3.3.5.

A Large Parameter

The saddle-point distribution function is asymptotic if the cumulant generating function depends on some large value of λ such that

$$\frac{dz}{dw}\frac{w}{z} \sim 1 + \left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) w \qquad \text{as } \lambda \to \infty \qquad \text{uniformly for } w \in {}_t\phi_r^*$$

Under these circumstances we expect that

$$F(x) \sim \Phi(\hat{w}) - \phi(\hat{w}) \left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) \qquad \text{as } \lambda \to \infty$$

Furthermore, the saddle-point method provides a complete asymptotic expansion for the distribution function if the inverse Fourier transform is of the form of equation (2.12). The sample mean is an example of such a case.

3.3.5 Example: The Non-central Chi-squared Distribution Function

This example continues from the non-central chi-squared saddle-point density discussed in section 3.2.6. Consider a non-central chi-squared random variable with degree of freedom p and non-centrality parameter η. The cumulant generating function of such a random variable is given by $K(t) = \frac{\eta t}{1-2t} - \frac{p}{2}\log(1-2t)$. The saddle-point $\hat{t}$ is given by

$$\hat{t}(x) = \frac{2x - p - \sqrt{p^2 + 4\eta x}}{4x}$$

Let $r(x) = \frac{p+\sqrt{p^2+4\eta x}}{4x}$ then it follows from section 3.2.6 that

$$K(\hat{t}(x)) - \hat{t}(x)x = \frac{1-r(x)}{2}\left(\frac{\eta}{r(x)} - x\right) - \frac{p}{2}\log r(x)$$
$$K^{(2)}(\hat{t}(x)) = \frac{4\eta}{r^3(x)} + \frac{2p}{r^2(x)} = \frac{2}{r^2(x)}\left(\frac{2\eta}{r(x)} + p\right)$$

Therefore it follows that

$$\begin{aligned}\hat{w} &= sign\{\hat{t}(x)\}\sqrt{2}\sqrt{\hat{t}\,x - K(\hat{t})} \\ &= sign\{x - \eta - p\}\sqrt{p\,\log r(x) - (1-r(x))\left(\frac{\eta}{r(x)} - x\right)}\end{aligned}$$

and

$$\begin{aligned}\hat{u} &= \hat{t}(x)\sqrt{K^{(2)}(\hat{t}(x))} \\ &= \frac{1-r(x)}{\sqrt{2}\,r(x)}\sqrt{\frac{2\eta}{r(x)} + p}\end{aligned}$$

The non-central chi-squared saddle-point distribution function is given by

$$F(x) \approx \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{u}} - \frac{1}{\hat{w}}\right) \qquad \text{for } x \neq \eta + p$$

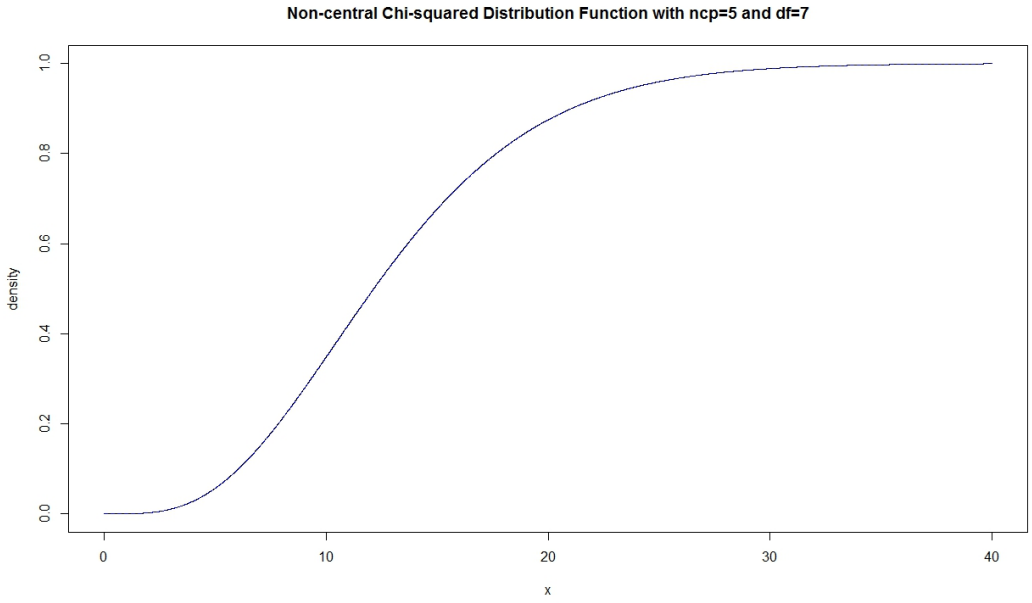


Figure 3.2: The exact distribution function is given by the solid line, and the saddle-point distribution function is given by the dashed blue line. The saddle-point distribution function lies on the exact distribution function and therefore provides an almost-perfect fit.

3.3.6 The Saddle-point Distribution Function of a Sample Mean

Consider a random sample $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables with the distribution function $F(x) = P(X \leq x)$ and cumulant generating function $K(t) = \log E\left[e^{tX}\right]$. Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ as the sample mean. The distribution function of the sample mean, say $F_n(\bar{x})$, is given by the inverse Fourier transformation:

$$F_n(\bar{x}) = \frac{1}{2\pi i}\lim_{r\to\infty}\int_{t\phi_r} e^{n(K(z)-z\bar{x})}\,\frac{dz}{z} \tag{3.11}$$

which is of the form (2.12) where $k = 1$ and $h(z) = K(z) - z\,x$. By applying the result given by (2.19), it follows that

$$\begin{aligned}
F_n(\bar{x}) &= \frac{1}{2\pi i}\int_{t\phi_r} e^{n(K(z)-z\bar{x})}\,\frac{dz}{z} \\
&= J_1\left(\sqrt{n}\hat{w}\right)\sum_{j=0}^{\infty}\frac{c_{2j}}{n^j} + J_0\left(\sqrt{n}\hat{w}\right)\sum_{j=0}^{\infty}\frac{c_{2j+1}}{n^{j+\frac{1}{2}}} \\
&= \Phi\left(\sqrt{n}\,\hat{w}\right)\left(c_0 + \frac{c_2}{n} + \frac{c_4}{n^2} + \ldots\right) + \frac{\phi\left(\sqrt{n}\,\hat{w}\right)}{\sqrt{n}}\left(c_1 + \frac{c_3}{n} + \frac{c_5}{n^2} + \ldots\right) \\
&= \left\{\Phi\left(\sqrt{n}\,\hat{w}\right) + \phi\left(\sqrt{n}\,\hat{w}\right)\left(\frac{1}{\hat{t}\sqrt{n\,K^{(2)}(\hat{t})}} - \frac{1}{\sqrt{n}\,\hat{w}}\right)\right\}\left(1 + O\left(n^{-1}\right)\right)
\end{aligned} \tag{3.12}$$

where $c_0, c_1, c_2, \ldots$ are real constants depending on $\hat{t}$ with $c_0 = 1$ and $c_1 = \frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}$.

The asymptotic expansion for the distribution function of a sample mean given by equation (3.12) is originally derived by Lugannani & Rice (1980). Daniels (1987) proves that the formula of Lugannani & Rice (1980) can be derived using a "less elaborate" method of Bleistein (1966).

3.4 Summary

3.4.1 Introduction

This section aims to capture the key points discussed in this chapter. The discussion of these key points are categorised according to the section in which they first appear, since many points appear in multiple sections.

3.4.2 The Saddle-point Density Function

An adjusted saddle-point method of Flajolet & Sedgewick (2009) is applied to the complex integral (3.1) to obtain the saddle-point density function given by equation (3.3) by choosing $h(z) = K(z) - z\,x$. The only condition that needs to be satisfied is that the quadratic approximation

$$K(\hat{t}+ir) - x\left(\hat{t}+ir\right) \approx K(\hat{t}) - \hat{t}\,x - \frac{1}{2}K^{(2)}(\hat{t})\,r^2$$

is uniformly adequate for $-\infty < r < \infty$. If this condition holds then it follows that

$$\begin{aligned} f(x) &= \frac{1}{2\pi i}\int_{\hat{\phi}_r} e^{K(z)-z\,x}\,dz && \text{equation (3.1)} \\ &\approx \frac{1}{2\pi} e^{K(\hat{t})-\hat{t}\,x} \lim_{r\to\infty}\int_{\hat{\phi}_r} e^{\frac{1}{2}K^{(2)}(\hat{t})\,(z-\hat{t})^2}\,dz && \text{from the condition} \\ &= \frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{2\pi\,K^{(2)}(\hat{t})}} && \text{equation (3.3)} \end{aligned}$$

which is called the saddle-point density (Butler, 2007:3). The saddle-point $\hat{t}$ is defined as the solution to the equation

$$K^{(1)}(\hat{t}) = x \qquad \text{such that } \hat{t} \in (t_1, t_2)$$

where $(t_1, t_2) \subset \mathbb{R}$ is the interval on which the cumulant generating function $K(t)$ is defined. The saddle-point path is given by $\hat{\phi}_r(s) = \hat{t} + is$ for $-r \le s \le r$ and it follows the direction of steepest descent.

Renormalising the Saddle-point Densty Function

The normalised saddle-point density function $\hat{f}(x)$ corresponding to the cumulant generating function $K(t)$ is defined as

$$\hat{f}(x) = \frac{1}{c}\frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{K^{(2)}(\hat{t})}} \qquad \text{where} \qquad c = \int_{-\infty}^{\infty}\frac{e^{K(\hat{t})-\hat{t}\,x}}{\sqrt{K^{(2)}(\hat{t})}}\,dx = \int_{t_1}^{t_2}\frac{e^{K(t)-t\,K^{(1)}(t)}}{\sqrt{K^{(2)}(t)}}\,dt$$

given that $K(t)$ is defined on the interval $(t_1, t_2) \subset \mathbb{R}$. The renormalised saddle-point density function is a real density function.

The integral c can be calculated using numerical methods. The function integrate in the programming language **R** is one such numerical method.

3.4.3 The Saddle-point Distribution Function

The saddle-point method which compensates for a simple pole at the origin is adjusted so that it can be applied the complex integral (3.6) to obtain the saddle-point distribution function given by equation (3.10) by choosing $h(z) = K(z) - z\,x$. The only condition that needs to be satisfied is that the linear approximation

$$\frac{dz}{dw}\,\frac{w}{z} \approx 1 + \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) w$$

is uniformly adequate for all w on the path ${}_t\phi_r(s) = t + is$ for $-r \leq s \leq r$ where $t \in (t_1, t_2)$ and $\hat{w}$ is defined as $\hat{w} = sign\{\hat{t}\}\,\sqrt{2}\sqrt{\hat{t}\,x - K(\hat{t})}$. If this condition holds then it follows that

$$\begin{aligned}
1 - F(x) &= \frac{1}{2\pi i} \lim_{r\to\infty} \int_{{}_t\phi_r} e^{K(z)-zx}\,\frac{dz}{z} && \text{equation (3.6)} \\
&= \frac{1}{2\pi i} \lim_{r\to\infty} \int_{{}_t\phi_r} e^{\frac{1}{2}w^2 - \hat{w}\,w}\,\frac{dz}{dw}\,\frac{dw}{z} \\
&\approx \frac{1}{2\pi i} \lim_{r\to\infty} \left\{ \int_{{}_t\phi_r} e^{\frac{1}{2}w^2 - \hat{w}\,w}\,\frac{dw}{w} + \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) \int_{{}_t\phi_r} e^{\frac{1}{2}w^2 - \hat{w}\,w}\,dw \right\} && \text{from the condition} \\
&= 1 - \Phi(\hat{w}) + \phi(\hat{w}) \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) && \text{equation (3.9)}
\end{aligned}$$

It furthermore follows from Butler (2007:69) that

$$\lim_{\hat{t}\to 0} \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) = \frac{1}{6}\,\frac{K^{(3)}(0)}{\left(K^{(2)}(0)\right)^{\frac{3}{2}}}$$

Therefore

$$F(x) \approx \begin{cases} \Phi(\hat{w}) - \phi(\hat{w}) \left(\frac{1}{\hat{t}\,\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) & \text{if } x \neq E[X] \\ \frac{1}{2} + \frac{1}{6\sqrt{2\pi}}\,\frac{K^{(3)}(0)}{\left(K^{(2)}(0)\right)^{\frac{3}{2}}} & \text{if } x = E[X] \end{cases} \qquad \text{equation (3.10)}$$

which is called the saddle-point distribution function (Butler, 2007:12).

Chapter 4

Applications to the Hill Estimator

4.1 Introduction

In this chapter the saddle-point density and distribution functions of the Hill estimator are investigated. The Hill estimator is one of the most popular estimators of the extreme value index (EVI) of a heavy-tailed distribution in extreme value theory (EVT). It is a measure of the heaviness of the underlying distribution's tail. The larger the EVI, the more prone the distribution is to exhibit extreme values. EVT is the statistical theory relating to extreme tail probabilities and quantiles. The first section in this chapter serves as an introduction to EVT and the EVI.

The Hill estimator is one of the earliest, and still one of the most popular, estimators of the EVI. If the tail of a distribution is assumed to be approximately Pareto, then the Hill estimator is the maximum likelihood estimator of the extreme value index. This Pareto approximation is, however, inaccurate and quite crude, and the Hill estimator itself is biased and unstable. However, the Hill estimator does have favourable asymptotic properties which simplifies inference. The Hill estimator and its properties are discussed in the third section of this chapter. This discussion serves as background to the saddle-point approximation of the Hill estimator.

The saddle-point density and distribution functions of the Hill estimator are derived in the fourth section of this chapter. The saddle-point approximation relies on a second-order approximation regarding the distribution of the Hill estimator. This second-order approximation together with its results are also discussed in the fourth section.

The final two sections of this chapter are dedicated to illustrating the accuracy and behaviour of the saddle-point density and distribution functions of the Hill estimator. The accuracy of the saddle-point approximation is compared to the Edgeworth expansion of Cuntz, Haeusler & Segers (2003) in the penultimate section. Furthermore, in the final section, the behaviour of the saddle-point density and distribution functions is investigated when the parameters are estimated using small samples.

4.2 Extreme Value Theory

4.2.1 Introduction

Extreme value theory (EVT) is the field in statistics that studies extreme quantiles and tail probabilities by relying on asymptotic theory. The two main asymptotic theorems in extreme value theory, namely the Fisher-Tippett-Gnedenko Theorem and the Pickands-Balkema-de Haan Theorem, give rise to the two most popular methods of modelling extreme quantiles and tail probabilities, which are the method of block maxima and the peaks-over-threshold method respectively. The asymptotic theory and methodology behind each method are briefly discussed in this chapter.

The extreme value index (EVI), denoted γ, is the central parameter in EVT and is a measure of the extremity of the underlying distribution. The EVI is the sole parameter to feature in the two main asymptotic theorems in EVT.

This chapter has three parts. Firstly the notation used in EVT is defined. Secondly the two main asymptotic theories in EVT are given and discussed. Thirdly the EVI and its role in EVT is discussed.

4.2.2 Notation

Consider a random sample $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables with cumulative distribution function F. The sample quantiles (or order statistics) are denoted by $X_{1,n} < X_{2,n} < \cdots < X_{n,n}$. The tail quantile function is defined as

$$T(x) = F^{-1}\left(1 - \frac{1}{x}\right) = \inf\left\{y : 1 - F(y) \leq \frac{1}{x}\right\}$$

The tail quantile function T plays a central role in EVT.

Independent standard exponential random variables are denoted as $E_1, E_2, \ldots, E_n$ and their corresponding sample quantiles (or order statistics) are denoted as $E_{1,n} < E_{2,n} < \cdots < E_{n,n}$. Renyi's representation of standard exponential order statistics yields

$$j\left(E_{n-j+1,n} - E_{n-j,n}\right) \stackrel{D}{=} E_j \qquad \text{for} \qquad j = 1, 2, \ldots, n$$

Independent standard uniform random variables are denoted as $U_1, U_2, \ldots, U_n$ and their corresponding sample quantiles (or order statistics) are denoted as $U_{1,n} < U_{2,n} < \cdots < U_{n,n}$.

4.2.3 Asymptotic Theorems

There are two main asymptotic theorems in EVT: the Fisher-Tippett-Gnedenko Theorem and Pickands-Balkema-de Haan Theorem. The former relates to the asymptotic distribution of a sample maximum and the latter relates to the asymptotic distribution of an exceedence. These two theorems are discussed in this section as background to the Hill estimator and its saddle-point density and distribution functions.

The Fisher-Tippett-Gnedenko Theorem is first devised by Fisher & Tippett (1928) and partly proven by Gnedenko (1943). The theorem solves the extremal limit problem (Beirlant et al., 2004:46) - if the sample maximum can be standardised so that it has a non-degenerate asymptotic distribution function G, then G has one of three forms: it is either of Frechet,

Weibull or Gumbel type, depending on the value of EVI.

The Fisher-Tippett-Gnedenko Theorem

Suppose that there exist sequences $\{a_n\}$ (which is positive) and $\{b_n\}$ such that

$$P\left(a_n^{-1}\left(X_{n,n}-b_n\right)\leq z\right)\to G(z) \qquad \text{as } n\to\infty$$

where G is a non-degenerate distribution function. Then $G\in\{G_\gamma\}$ where

$$G_\gamma(z)=\begin{cases} e^{-(1+\gamma\, z)^{-\frac{1}{\gamma}}} & \text{if } \gamma>0\\ e^{-e^{-z}} & \text{if } \gamma=0\\ e^{-(1+\gamma\, z)^{-\frac{1}{\gamma}}} & \text{if } \gamma<0 \end{cases}$$

The parameter γ is called the extreme value index.

Gnedenko (1943) solves the domain of attraction problem (Beirlant et al. 2004:46) - if the condition C_γ(a) is satisfied then a sample maximum with distribution function F^n can be standardised so that it has a non-degenerate asymptotic distribution function G.

The Condition $C_\gamma(a)$

For a distribution function F the condition $C_\gamma(a)$ holds if a positive function $a(x)$ and a real constant γ exist such that

$$C_\gamma(a): \qquad \frac{T(xu)-T(x)}{a(x)}\to h_\gamma(u) \qquad \text{as } x\to\infty \tag{4.1}$$

where $h_\gamma(u)=\begin{cases}\frac{u^\gamma-1}{\gamma} & \text{if } \gamma\neq 0\\ \log u & \text{if } \gamma=0\end{cases}$

If the tail quantile function $T(x)$ satisfies the condition $C_\gamma(a)$, then it is said that the distribution function F lies in the domain of attraction of G_γ. This is denoted by $F\in D(G_\gamma)$.

The Fisher-Tippett-Gnedenko Theorem holds if and only if the condition $C_\gamma(a)$ is satisfied. As a direct consequence it follows that a distribution function F is in the domain of attraction of G_γ with $\gamma>0$ if and only if

$$T(x)=x^\gamma\, l_T(x) \qquad \text{or equivalently} \qquad 1-F(x)=x^{-\frac{1}{\gamma}}\, l_F(x) \tag{4.2}$$

for some slowly varying functions (Beirlant et al. 2004:49) l_T and l_F which relate to one another via the so-called de Bruijn conjugate (Beirlant et al. 2004:57-58). A distribution function F which satisfies equations (4.3) is said to be Pareto-type. The equations in (4.3) are of a semi-parametric form because each consists of a parametric component and a non-parametric component.

In order to model the tail quantile function $T(x)$ it is necessary to introduce some structure on the slowly varying part $l_T(x)$. The first order approximation of $T(x)$ effectively ignores $l_F(x)$ for large x so that the tail of the distribution is approximately Pareto, ie $T(x)\approx x^\gamma$ for large x.

The second-order approximation of $T(x)$ entails that $\log l_T(x)$ satisfies $C_\rho(b)$ where $\rho < 0$ and $b(x)$ is regularly varying with index ρ (Beirlant et al. 2004:49). The second-order approximation and its implications are further discussed in section 4.4.2.

The Generalised Extreme Value Distribution

For a given sample of size n the distribution of the maximum can be approximated by

$$F^n(x) = P\left(X_{n,n} \leq x\right) \approx G_\gamma\left(\frac{x-\mu}{\sigma}\right) = G_{\mu,\sigma,\gamma}(x) \qquad \text{for large } n$$

The distribution function $G_{\mu,\sigma,\gamma}(x)$ is called the generalised extreme value (GEV) distribution function and is given by

$$G_{\mu,\sigma,\gamma}(x) = G_\gamma\left(\frac{x-\mu}{\sigma}\right) = \exp\left\{-\left(1+\frac{\gamma}{\sigma}\left(x-\mu\right)\right)^{-\frac{1}{\gamma}}\right\} \tag{4.3}$$

where the parameters μ and σ are the location and scale parameters and γ is the EVI. The parameters μ, σ and γ can be fitted using the method of block maxima.

The Method of Block Maxima

The method of block maxima employs the Fisher-Tippett Theorem as a means to model the tail of the underlying distribution. The method entails dividing the observations into m blocks, each of size r. The maximum of each block is called a block maximum, and collectively they are known as the block maxima. The GEV distribution is fitted using maximum likelihood where the block maxima are used as the observations.

The Pickands-Balkema-de Haan Theorem is devised and proven by Pickands (1975) and Balkema & de Haan (1974). The theorem provides an alternative and equivalent condition to $C_\gamma(a)$ which is denoted by $C^*_\gamma(b)$.

Pickands-Balkema-de Haan Theorem

The distribution function F is in the domain of attraction of G_γ if and only if there exists some auxiliary function $b(t)$ such that

$$C^*_\gamma(b): \qquad \frac{1-F(t+b(t)\,y)}{1-F(t)} \to (1+\gamma\,y)^{-\frac{1}{\gamma}} \qquad \text{as } t \to \infty \tag{4.4}$$

The Generalised Parto Distribution

The survival function of an exceedence $Y = X - t | X > t$ with threshold t can be approximated by

$$P(Y > y) = \frac{1-F(y+t)}{1-F(t)} \approx \left(1+\frac{\gamma}{\sigma}\,y\right)^{-\frac{1}{\gamma}} = 1 - H_{\sigma,\gamma}(y) \qquad \text{for large } t$$

where $\sigma = b(t)$ (Beirlant et al. 2004:147). The distribution function $H_{\sigma,\gamma}(x)$ is called the generalised Pareto distribution (GPD) function where σ is the scale parameter and γ is the EVI. The parameters σ and γ can be fitted using the peaks-over-threshold method.

The Peaks-over-Threshold Method

The peaks-over-threshold method employs the Pickands-Balkema-de Haan Theorem as a means to model the tail of the underlying distribution. The method entails choosing a high threshold and considering the observations that exceed this threshold. The amount by which an observation exceeds the threshold is called an exceedence, and collectively they are known as the exceedences. The GPD is fitted using maximum likelihood where the exceedences are used as the observations.

4.2.4 The Extreme Value Index

The EVI γ is the sole parameter to feature in the asymptotic distribution of both the standardised maximum and the standardised exceedence. The parameter γ determines the heaviness of the tail of the distribution F - a larger value of γ implies a heavier tail.

In EVT, a distribution is categorised according to its EVI, which indicates the heaviness of the distribution's tail:

If $\gamma < 0$ then the distribution falls in the Weibull class. A distribution in this class has a finite maximum. This finite maximum can be estimated using the peaks-over-threshold method.

If $\gamma = 0$ then the distribution falls in the Gumbel class. A distribution in this class has a survival function which decreases exponentially.

If $\gamma > 0$ then the distribution falls in the Frechet-Pareto class. A distribution in this class has a survival function which decreases at a polynomial speed. If $\gamma > \frac{1}{2}$ then the variance is infinite and if $\gamma > 1$ then the mean of the distribution does not exist. (Beirlant et al., 2004:58).

In this paper we are principly interested in the Frechet-Pareto class of distributions since the Hill estimator is one of the most popular estimators of the EVI for this class. Distributions from the Frechet-Pareto class feature in a variety of settings, specifically in finance and insurance. Returns on financial instruments, such as stocks and indices, frequently have heavy tailed distributions. The size of insurance claims also frequently have heavy tailed distributions. Insurance claims are often related to damage from natural disasters which are events that are both rare and of a very large magnitude.

4.3 The Hill Estimator

4.3.1 Introduction

The Hill estimator is one of the most popular estimators of a positive EVI. The Hill estimator is biased and rather unstable. The accuracy of the Hill estimator relies on the tail of the distribution being approximately Pareto. This first-order approximation, which is briefly discussed in section 4.2.3, does not hold in general. However, the asymptotic properties of the Hill estimator makes inference using the estimator quite straightforward, given that the asymptotic conditions are met.

This section serves as an introduction to the Hill estimator and has two parts. Firstly, the Hill estimator is defined. Secondly, the properties of the Hill estimator under the first-order approximation are discussed.

4.3.2 Definition

The Hill estimator is given by

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j$$

where $Z_j = j\,(\log X_{n-j+1,n} - \log X_{n-j,n})$ for $j = 1, 2, \ldots, k$ (Beirlant et al. 2004:101-103). The integer k can be regarded as the tuning parameter of the Hill estimator, and needs to be chosen. The Hill estimator is often plotted against k to yield the Hill plot (Beirlant et al. 2004:104).

The choice of the integer k affects the bias and variance of the Hill estimator. The smaller the value of k, the larger the variance and the smaller the bias of the Hill estimator. The larger the value of k, the smaller the variance and the larger the bias of the Hill estimator. This phenomenon is called the bias-variance tradeoff. Methods for chosing k within this context are discussed in Beirlant et al. (2004), section 4.7.

4.3.3 The Hill Estimator as a Maximum Likelihood Estimator

The Hill estimator is a maximum likelihood estimator of a positive EVI under the first-order approximation of $T(x)$. This result can be derived in two ways: by either considering the tail quantile function $T(x) = x^{\gamma}\, l_T(x)$ or the survival function $1 - F(x) = x^{-\frac{1}{\gamma}}\, l_F(x)$. The semi-parametric form of these two functions is implied by (4.3).

Maximum Likelihood Estimator based on The Tail Quantile Function

If F is in the domain of attraction of G_γ and $\gamma > 0$ then it follows from (4.3) that $T(x) = x^{\gamma}\, l_T(x)$. The first-order approximation of $T(x)$ is given by

$$T(x) \approx x^{\gamma} \qquad \text{for large } x \tag{4.5}$$

The first-order approximation (4.6) implies that

$$\begin{aligned} \log X_{j,n} &= \log F^{-1}(U_{j,n}) \\ &= \log T\left(\frac{1}{1-U_{j,n}}\right) \\ &\approx -\gamma \log(1-U_{j,n}) \\ &= \gamma\, E_{j,n} \end{aligned}$$

for $j = 1, 2, \ldots, k$ where k is small and n is large. Furthermore,

$$\begin{aligned} Z_j &= j\ (\log X_{n-j+1,n} - \log X_{n-j,n}) \\ &\approx j\,(\gamma\, E_{n-j+1,n} - \gamma\, E_{n-j,n}) \\ &= \gamma\, j\ (E_{n-j+1,n} - E_{n-j,n}) \\ &= \gamma\, E_j \end{aligned}$$

for $j = 1, 2, \ldots, k$ where k is small and n is large. The density function of Z_j can therefore approximated by

$$f_{Z_j}(z) \approx \frac{1}{\gamma}\, e^{-\frac{1}{\gamma} z} \qquad \text{for } j = 1, 2, \ldots, k$$

It follows that the maximum likelihood estimator of γ can be approximated by

$$\hat{\gamma}_k = \frac{1}{k} \sum_{j=1}^{k} Z_j = H_{k,n}$$

which is the Hill estimator.

Maximum Likelihood Estimator based on $F(x)$ Assumption

If F is in the domain of attraction of G_γ and $\gamma > 0$ then it follows from (4.3) that

$$\frac{1 - F(xt)}{1 - F(t)} = x^{-\frac{1}{\gamma}} \frac{l_F(xt)}{l_F(x)} \to x^{-\frac{1}{\gamma}} \qquad \text{as } t \to \infty$$

(Beirlant, et al. 2004:57). Define $Y = \frac{X}{t} \big| X > t$ as the relative exceedence and $n_t = \sum_{i=1}^{n} Ind\{X_i > t\}$ as the number of exceedences. It follows that

$$P(Y > y) \approx y^{-\frac{1}{\gamma}} \qquad \text{and} \qquad f_Y(y) \approx \frac{1}{\gamma}\, y^{-\frac{1}{\gamma}} \qquad \text{for a large threshold } t$$

The maximum likelihood estimator of γ can therefore be approximated by

$$\hat{\gamma}_{n_t} = \frac{1}{n_t} \sum_{i=1}^{n_t} \log Y_i$$

By choosing $n_t = k$ implies that $t = X_{n-k,n}$ which yields the Hill estimator

$$\hat{\gamma}_k = \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n}$$

4.3.4 Remarks

The aim of this section is to define the Hill estimator and to provide some insight into the characteristics of the Hill estimator using the first-order approximation. The first-order approximation, however, is not accurate in general. This gives rise to the second-order approximation of $T(x)$ which is discussed in the next section. The second-order approximation yields accurate results when used to approximate the distribution of the Hill estimator.

4.4 Approximations of the Distribution of the Hill Estimator

4.4.1 Introduction

The aim of this chapter is to derive saddle-point density and distribution functions for the second-order approximation of the Hill estimator which is of the form

$$H_{k,n} \approx \frac{1}{k} \sum_{j=1}^{k} \mu_j \, E_j \tag{4.6}$$

Equation (4.7) is the mean of k independent and non-identically distributed exponential random variables with means $\mu_1, \mu_2, \ldots, \mu_k$. The saddle-point density and distribution functions of (4.7) serve as a second-order approximation of the density and distribution functions of the Hill estimator respectively.

It is possible to derive the exact distribution of the mean of k independent and non-identically distributed exponential random variables by using convolutions (Akkouchi, 2009). However, in the case of the second-order approximation of the Hill estimator given by equation (4.7), it turns out that the exact distribution of Akkouchi (2008) is unstable.

The accuracy of the saddle-point density and distribution functions of the Hill estimator is illustrated in the example at the end of this section. The saddle-point density and distribution functions are compared to the Edgeworth expansion density and distribution functions of Cuntz, Haeusler & Segers (2003:3) which are based on the same second-order approximation.

4.4.2 The Second-Order Approximation

If F is in the domain of attraction of G_γ and $\gamma > 0$ then $T(x) = x^\gamma \, l_T(x)$ where l_T is a slowly-varying function. This is equivalent to assuming that the distribution function F is of Pareto-type: $1 - F(x) = x^{-\frac{1}{\gamma}} \, l_F(x)$ where l_F is slowly-varying and $\gamma > 0$. It follows that for $j = 1, 2, \ldots, n$:

$$\begin{aligned}
\log X_{j,n} &= \log F^{-1}(U_{j,n}) \\
&= \log T\left(\frac{1}{1 - U_{j,n}}\right) \\
&= -\gamma \, \log(1 - U_{j,n}) + \log l_T\left(\frac{1}{1 - U_{j,n}}\right) \\
&= \gamma \, E_{j,n} + \epsilon_{j,n}
\end{aligned}$$

where $\epsilon_{j,n} = \log l_T\left(\frac{1}{1-U_{j,n}}\right)$ is a random variable. Furthermore,

$$\begin{aligned}
Z_j &= j \, (\log X_{n-j+1,n} - \log X_{n-j,n}) \\
&= j \, (\gamma \, E_{n-j+1,n} + \epsilon_{n-j+1,n} - \gamma \, E_{n-j,n} - \epsilon_{n-j,n}) \\
&= \gamma \, j \, (E_{n-j+1,n} - E_{n-j,n}) + j \, (\epsilon_{n-j+1,n} - \epsilon_{n-j,n}) \\
&= \gamma \, E_j + j \, \Delta\epsilon_j
\end{aligned}$$

where $\Delta\epsilon_j = \epsilon_{n-j+1,n} - \epsilon_{n-j,n}$.

The Second Order Approximation

Assume that $\log l_T$ satisfies $C_\rho(b)$ where $\rho < 0$ and $b(x)$ is regularly varying with index ρ. Then we have the following two approximations:

1 $\log \frac{l_T(xu)}{l_T(x)} \sim b(x)\, h_\rho(u)$ as $x \to \infty$, therefore

$$\begin{aligned} j\,\Delta\epsilon_j &= j\,\log l_T\left(\frac{1}{1-U_{j,n}}\right) - \log l_T\left(\frac{1}{1-U_{j,n}}\right) \\ &= j\,\log \frac{l_T\left(e^{E_{n-j+1,n}}\right)}{l_T\left(e^{E_{n-j,n}}\right)} \\ &= j\,\log \frac{l_T\left(e^{E_{n-j+1,n}-E_{n-j,n}}\,e^{E_{n-j,n}}\right)}{l_T\left(e^{E_{n-j,n}}\right)} \\ &= j\,\log \frac{l_T\left(e^{\frac{1}{j}E_j}\,e^{E_{n-j,n}}\right)}{l_T\left(e^{E_{n-j,n}}\right)} \\ &\approx j\,b\left(e^{E_{n-j,n}}\right)\,\frac{e^{\frac{\rho}{j}E_j}-1}{\rho} \qquad \text{for small } j \text{ relative to large } n \\ &= b\left(e^{E_{n-j,n}}\right)\,W_j \end{aligned}$$

where $W_j = -\frac{j}{\rho}\left(1 - e^{\frac{\rho}{j}E_j}\right)$ has the distribution function

$$P\left(W_j \le w\right) = 1 - \left(1 + \frac{\rho}{j}\,y\right)^{-\frac{j}{\rho}} \approx 1 - e^{-w} \qquad \text{for large } j \text{ or } \rho \text{ close to } 0$$

W_j is approximately standard exponential for large j or ρ close to 0.

2 $\frac{b(xu)}{b(x)} \to u^\rho$ as $x \to \infty$ and $E_{n-j,n} \approx \log \frac{n}{j}$, therefore

$$b\left(e^{E_{n-j,n}}\right) \approx b\left(\frac{n}{j}\right) \sim \left(\frac{j}{k}\right)^{-\rho} b\left(\frac{n}{k}\right) \qquad \text{for large } n \text{ and small } k$$

By combining these two results, it follows that that

$$\begin{aligned} Z_j &= \gamma\,E_j + j\,\Delta\epsilon_j \\ &\approx \gamma\,E_j + b\left(e^{E_{n-j,n}}\right)\,W_j \\ &\approx \gamma E_j + b_{n,k}\left(\frac{j}{k}\right)^{-\rho} W_j \qquad \text{for } j = 1, 2, \ldots, k \end{aligned}$$

where k is small, n is large and $b_{n,k} = b\left(\frac{n}{k}\right)$.

By further assuming that W_j is approximately standard exponential for large j or ρ close to 0, it follows that

$$Z_j \overset{D}{\approx} \left(\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}\right) E_j \qquad \text{for} \qquad 1, 2, \ldots, k \tag{4.7}$$

where k is small, n is large, $\gamma > 0, b_{n,k} > 0$ and $\rho < 0$. Equation (4.8) is the second-order approximation of Z_j.

4.4.3 The Asymptotic Distribution of the Hill Estimator

Assume that $\log l_T$ satisfies $C_\rho(b)$ where $\rho < 0$ and $b(x)$ is regularly varying with index ρ such that $\sqrt{k}\, b\left(\frac{n}{k}\right) \to c$ for some $c \in [0, \infty)$ as $n \to \infty$ then

$$\frac{\sqrt{k}\,(H_{k,n} - \gamma)}{\gamma} \xrightarrow{D} \begin{cases} N(0,1) & \text{if } c = 0 \\ N\left(\frac{c}{\gamma\,(1-\rho)}\right) & \text{if } c > 0 \end{cases} \qquad \text{as } n \to \infty$$

(Cuntz, Haeusler & Segers, 2003:2).

4.4.4 The Edgeworth Expansion of the Hill Estimator

Cuntz, Haeusler & Segers (2003) derive an Edgeworth expansion of the Hill estimator under the second order approximation (4.8). This Edgeworth expansion is given as follows:

$$P\left(\frac{\sqrt{k}\,(H_{n,k} - \gamma)}{\gamma} \leq x\right) = \Phi(x) + \phi(x)\left\{\frac{1-x^2}{3\sqrt{k}} - \frac{\sqrt{k}\, b_{n,k}}{\gamma(1-\rho)}\right\} + O\left(\frac{1}{k}\right) + o\left(\frac{\sqrt{k}}{b_{n,k}}\right) \tag{4.8}$$

The corresponding Edgeworth expansion density function is given by

$$f_{H_{n,k}}(x) \sim \frac{\sqrt{k}}{\gamma\sqrt{2\pi}}\, e^{-\frac{k}{2\gamma^2}\left(x-\gamma-\frac{b_{n,k}}{1-\rho}\right)^2}\left\{1 + \frac{x-\gamma}{\gamma}\left(\frac{k\, b_{n,k}}{\gamma\,(1-\rho)} - 1\right) + \frac{k\,(x-\gamma)^3}{3\,\gamma^3}\right\} \tag{4.9}$$

This Edgeworth expansion of Cuntz, Haeusler & Segers (2003) is an improvement on the Edgeworth expansion of Cheng & Pan (1998).

4.4.5 The Second Order Approximation of the Hill Estimator

The Hill estimator is defined in section 4.3.2 as

$$H_{n,k} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j$$

Define $\mu_j = \gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}$ for $j = 1, 2, \ldots, k$. It follows from (4.8) that the second-order approximation of the Hill estimator is given by

$$\tilde{H}_{n,k} = \frac{1}{k}\sum_{j=1}^{k} \mu_j E_j$$

where $E_1, E_2, \ldots, E_k$ are independent standard exponential random variables.

The Exact Distribution of $\tilde{H}_{n,k}$

The second-degree approximation of the Hill estimator, denoted $\tilde{H}_{k,n}$, is the sample mean of k independent exponential random variables with means $\mu_1, \mu_2, \ldots, \mu_k$ respectively. Akkouchi (2008) derives the exact distribution of the sample mean of independent exponential random variables with different means using mathematical induction on convolutions. It follows from Akkouchi (2008) that the density function of $\tilde{H}_{k,n}$ is given by

$$f_{\tilde{H}_{n,k}}(x) = k\sum_{j=1}^{k} \frac{\mu_j\, e^{-\frac{k}{\mu_j}x}}{\prod_{i\neq j}^{k}(\mu_j - \mu_i)} \tag{4.10}$$

By integrating the density function $f_{\tilde{H}_{k,n}}(x)$ it follows that the corresponding distribution function is given by

$$F_{\tilde{H}_{n,k}}(x) = \sum_{j=1}^{k} \frac{\mu_j^2 \left(1 - e^{-\frac{k}{\mu_j} x}\right)}{\prod_{i \neq j}^{k} (\mu_j - \mu_i)}$$

From the definition of $\{\mu_j\}$ it follows that $\gamma < \mu_1 < \mu_2 < \cdots < \mu_k = \gamma + b_{n,k}$. The denominator of the density function and distribution function of $\tilde{H}_{k,n}$ is bounded as follows:

$$0 < \prod_{i \neq j}^{k} (\mu_j - \mu_i) < b_{n,k}^{k-1} \to 0 \qquad \text{as } k \to \infty \quad \text{for} \quad 0 < b_{n,k} < 1$$

The denominator decreases to 0 exponentially as k increases, given that $0 < b_{n,k} < 1$. Therefore the density and distribution functions of $H_{k,n}$ are unstable for large k and $b_{n,k}$ close to 0.

The Saddle-point Density and Distribution Functions of $\tilde{H}_{n,k}$

The saddle-point method can be used to approximate the density function and distribution function of $\tilde{H}_{k,n}$. The moment generating function of $\tilde{H}_{k,n}$ is given by

$$M(t) = E\left[e^{t\,\tilde{H}_{k,n}}\right] = E\left[e^{\frac{t}{k}(\mu_1 E_1 + \mu_2 E_2 + \cdots + \mu_k E_k)}\right] = \left(1 - \frac{\mu_1\, t}{k}\right)^{-1} \left(1 - \frac{\mu_2\, t}{k}\right)^{-1} \cdots \left(1 - \frac{\mu_3\, t}{k}\right)^{-1}$$

and the corresponding cumulant generating function is given by

$$K(t) = \log M(t) = -\sum_{j=1}^{k} \log\left(1 - \frac{\mu_j\, t}{k}\right)$$

From the definition of $\{\mu_j\}$ it follows that $\gamma < \mu_1 < \mu_2 < \cdots < \mu_k = \gamma + b_{n,k}$. The cumulant generating function has singularities at $t \in \left\{\frac{k}{\mu_1}, \frac{k}{\mu_2}, \ldots, \frac{k}{\mu_k}\right\}$ and is therefore defined on the interval $\left(-\infty, \frac{k}{\mu_k}\right)$. The saddle-point is given by the equation

$$\begin{aligned} K^{(1)}(\hat{t}) &= x \\ \sum_{j=1}^{k} \frac{\mu_j}{k - \mu_j\, \hat{t}} &= x \qquad \text{such that} \quad \hat{t} < \frac{k}{\mu_k} \end{aligned} \tag{4.11}$$

In order to determine the saddle-point $\hat{t}$, it is necessary to find its lower bound. By using (4.12) together with the fact that $\gamma < \mu_1 < \mu_2 < \cdots < \mu_k = \gamma + b_{n,k}$ it follows that $\hat{t}$ is bounded as follows

$$\begin{aligned} \frac{k}{\mu_k} - \frac{k}{x} < \hat{t} &< \min\left\{\frac{k}{\mu_k}, \frac{k}{\mu_1} - \frac{k}{x}\right\} \\ \frac{k}{\gamma + b_{n,k}} - \frac{k}{x} < \hat{t} &< \min\left\{\frac{k}{\gamma + b_{n,k}}, \frac{k}{\gamma} - \frac{k}{x}\right\} \end{aligned}$$

Equation (4.12) can be solved for a vector of values $\underline{x}$ given the vector of mean values $\underline{\mu}$ using the uniroot function in the programming language R.

The saddle-point density function of $\tilde{H}_{k,n}$ is given by

$$f_{\tilde{H}_{n,k}}(x) \approx \frac{e^{K(\hat{t})-\hat{t}\,x}}{2\pi\, K^{(2)}(\hat{t})} \tag{4.12}$$

and the saddle-point distribution function of $\tilde{H}_{k,n}$ is given by

$$F_{\tilde{H}_{n,k}}(x) \approx \begin{cases} \Phi(\hat{w}) - \phi(\hat{w}) \left(\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}} \right) & \text{if } x \neq \bar{\mu} \\ \frac{1}{2} + \frac{1}{6\sqrt{2\pi}} \frac{\sum_{j=1}^{k} \mu_j^3}{\left(\sum_{j=1}^{k} \mu_j^2\right)^{\frac{3}{2}}} & \text{if } x = \bar{\mu} \end{cases} \tag{4.13}$$

where $\hat{w} = sign\{\hat{t}\}\,\sqrt{2}\sqrt{\hat{t}\,x - K(\hat{t})}$ and $\bar{\mu} = \frac{1}{k}\sum_{j=1}^{k} \mu_j$. The saddle-point density and distribution functions of $\tilde{H}_{k,n}$ are used as second-order approximations of the density and distribution functions of the Hill estimator $H_{k,n}$. This is discussed in the following section.

4.4.6 The Saddle-point Density and Distribution Functions of the Hill Estimator

The saddle-point density and distribution functions given by (4.13) and (4.14) are approximations of the density and distribution functions of the Hill estimator $H_{k,n}$ respectively. This follows from two explicit approximations:

- Under the second-order approximation discussed in section 4.4.2 which results in equation (4.8), it follows that $H_{k,n} \overset{D}{\approx} \tilde{H}_{k,n}$.

- The density and distribution functions of $\tilde{H}_{k,n}$ are approximated by the saddle-point density and distribution functions respectively.

The saddle-point density and distribution functions are approximations and not necessarily asymptotic results. In order to establish the asymptotic nature of these two functions it is necessary to define the behaviour of k as $n \to \infty$ and, consequently, the behaviour of $b_{n,k} = b\left(\frac{n}{k}\right)$ as $n \to \infty$, and to see whether condition 2 of Flajolet & Sedgewick (2009:552) is satisfied. The asymptotic nature of the saddle-point density and distribution functions is not investigated further in this thesis.

4.5 The Accuracy of the Approximations of the Distribution of the Hill Estimator

4.5.1 Introduction

The aim of this section is to illustrate the accuracy of the approximations of the distribution of the Hill estimator. Two approximations of the distribution of the Hill estimator are considered, namely the Edgeworth expansion of Cuntz, Haeusler & Segers (2003) of section 4.4.4 and the saddle-point density and distribution functions of equations (4.13) and (4.14) respectively. These two approximations are illustrated for two heavy-tailed distributions, namely the Burr and Frechet distributions, by plotting them against the histogram and empirical distribution function of simulated Hill estimators respectively. The empirical distribution function and histogram of the Hill estimator for each distribution is based on 10^5 simulated Hill estimates.

4.5.2 The Hill Estimator of a Burr Random Sample

Consider the random sample $X_1, X_2, \ldots, X_n$ where each random variable has a $Burr\,(a, b)$ distribution. It follows that

$$l_T(x) = x^{-\frac{1}{ab}}\, T(x) = \left(1 - x^{-\frac{1}{a}}\right)^{\frac{1}{b}} \qquad \text{and} \qquad b(x) \sim \frac{1}{ab}\left(x^{\frac{1}{a}} - 1\right)^{-1}$$

from which it follows that

$$\gamma = \frac{1}{ab} \qquad \rho = -\frac{1}{a} \qquad b_{n,k} = \frac{1}{ab}\left(\left(\frac{n}{k}\right)^{\frac{1}{a}} - 1\right)^{-1}$$

The saddle-point density and distribution functions of the Hill estimator are illustrated by the following two graphs. The two shape parameters are chosen as $a = \frac{1}{2}$ and $b = \frac{2}{3}$ respectively. The sample size is chosen as 100 and the parameter k is chosen as 10.

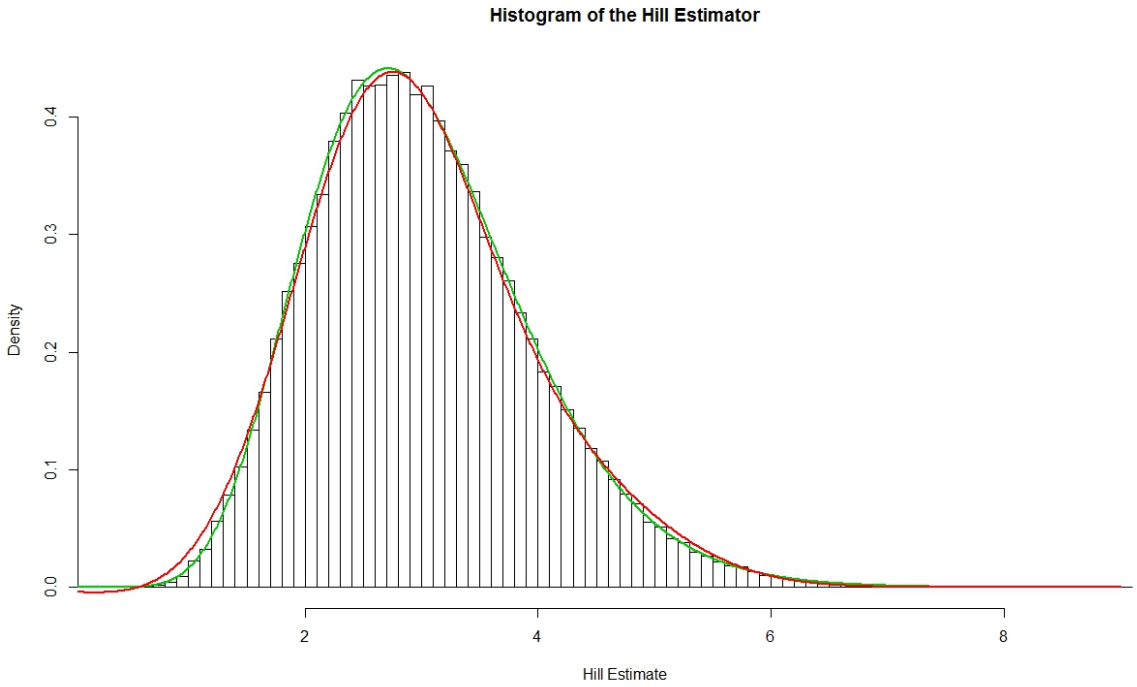


Figure 4.1: The green and red lines represent the saddle-point and Edgeworth density functions of the Hill estimator respectively.

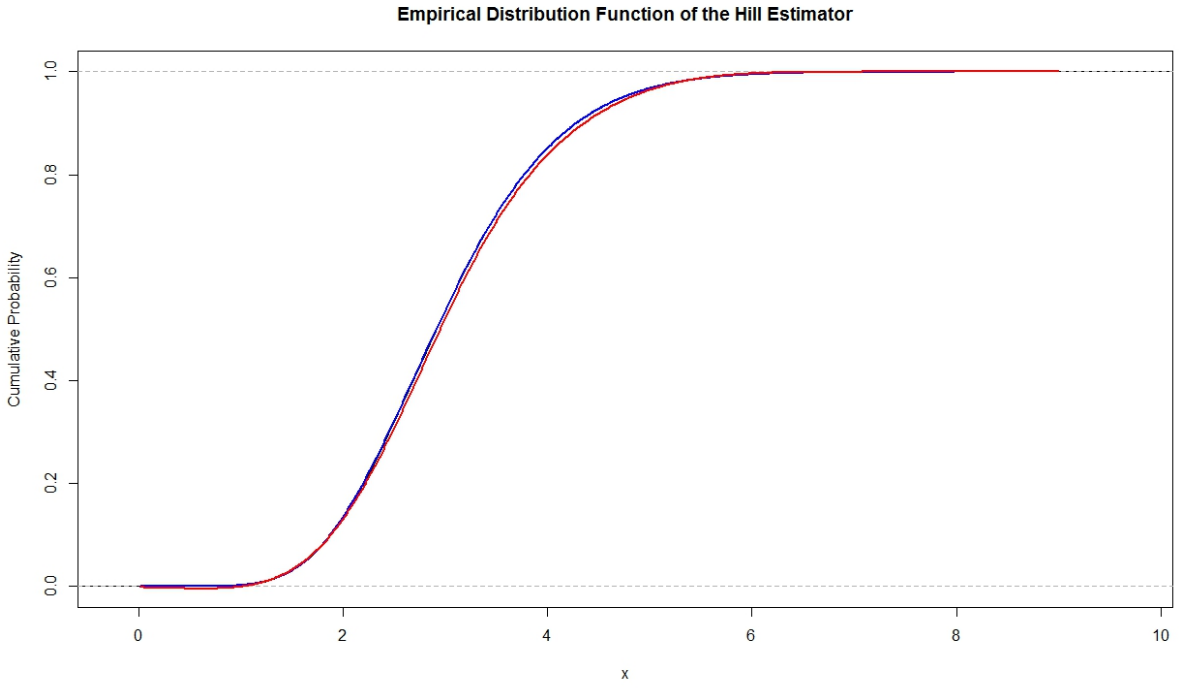


Figure 4.2: The blue and red lines represent the saddle-point and Edgeworth distribution functions of the Hill estimator respectively.

The graphs illustrate that both the saddle-point density and distribution functions are highly accurate approximations. Furthermore, the saddle-point density is a probability density function and the saddle-point distribution function is a cumulative distribution function.

The first-order Edgeworth expansion density function is negative for $x < \frac{1}{2}$ and is therefore

not a probability density function. This is reflected in the first order Edgeworth expansion distribution function which is also negative for $x < \frac{1}{2}$ and is therefore not a cumulative distribution function. The first-order Edgeworth expansion is only accurate over the central domain.

4.5.3 The Hill Estimator of a Frechet Random Sample

Consider the random sample $X_1, X_2, \ldots, X_n$ where each random variable has a $Frechet\,(a)$ distribution. It follows that

$$l_T(x) = x^{-\frac{1}{a}}\,T(x) = \left(-x\,\log\left(1-\frac{1}{x}\right)\right)^{-\frac{1}{a}} \qquad \text{and} \qquad b(x) \sim \frac{1}{2ax}$$

from which it follows that

$$\gamma = \frac{1}{a} \qquad \rho = -1 \qquad b_{n,k} = \frac{k}{2an}$$

The saddle-point density function and distribution functions of the Hill estimator are illustrated by the following two graphs. The shape parameter is chosen as $a = \frac{1}{3}$. The sample size is chosen as 100 and the parameter k is chosen as 10.

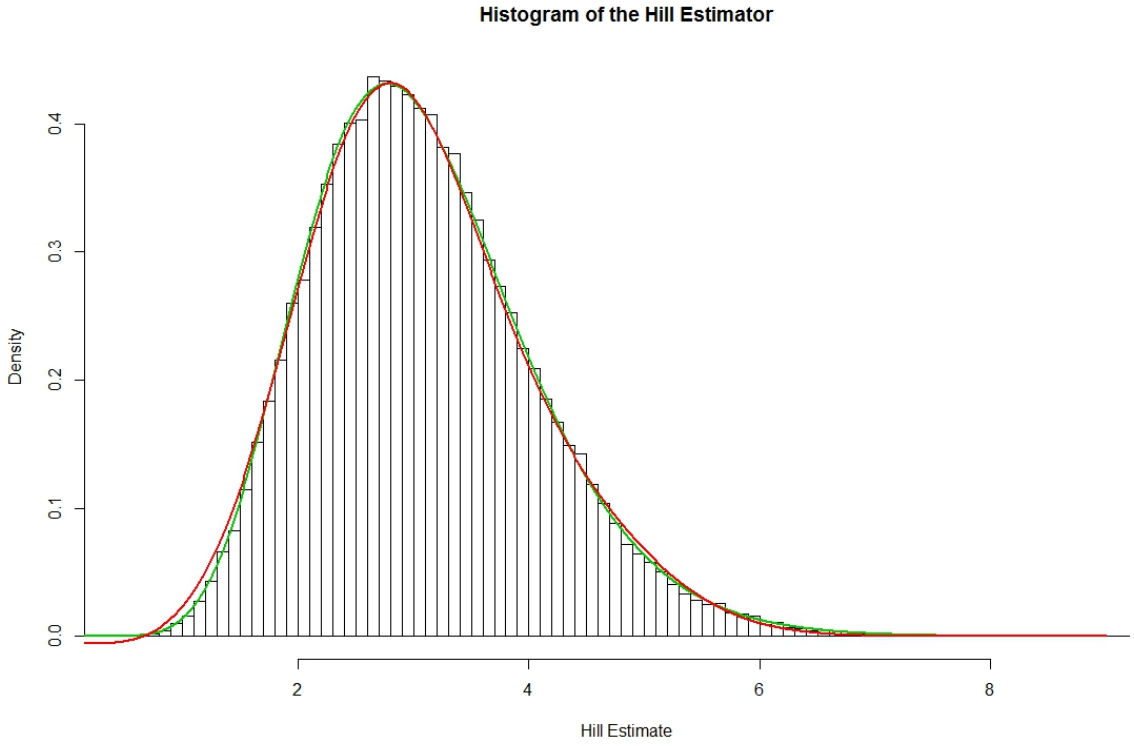


Figure 4.3: The green and red lines represent the saddle-point and Edgeworth density functions of the Hill estimator respectively.

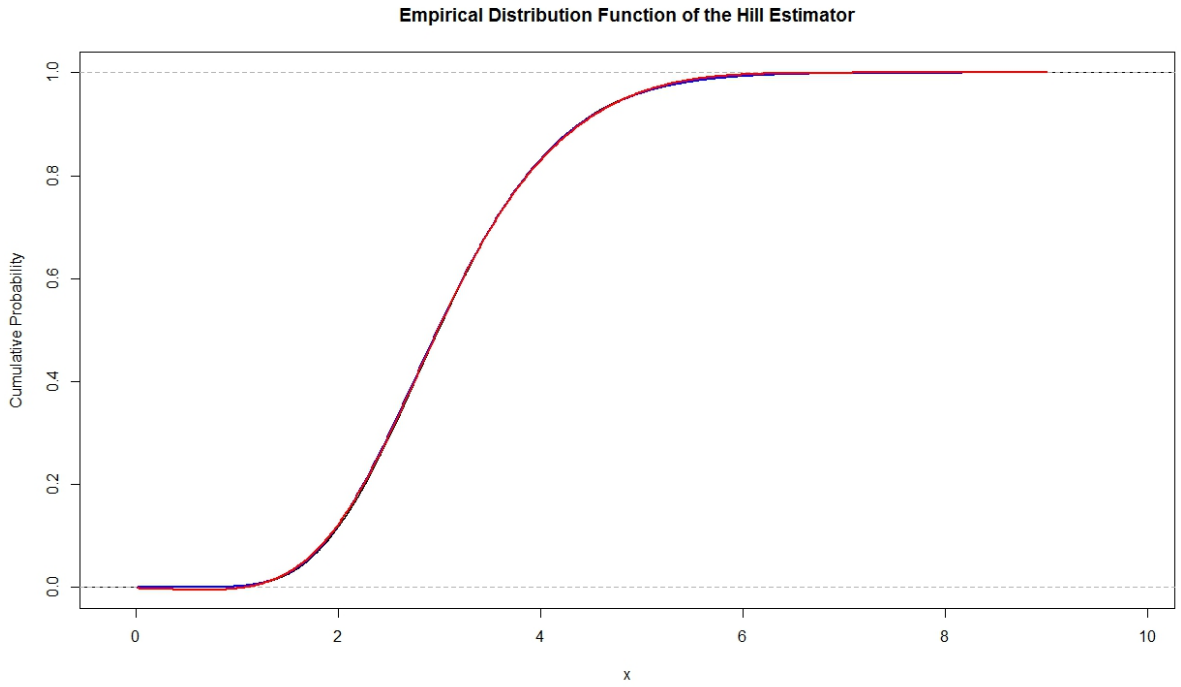


Figure 4.4: The blue and red lines represent the saddle-point and Edgeworth distribution functions of the Hill estimator respectively.

The graphs illustrate that both the saddle-point density and distribution functions are highly accurate approximations. Furthermore, the saddle-point density is a probability density function and the saddle-point distribution function is a cumulative distribution function.

The first-order Edgeworth expansion density function is negative for $x < \frac{1}{2}$ and is therefore not a probability density function. This is reflected in the first order Edgeworth expansion distribution function which is also negative for $x < \frac{1}{2}$ and is therefore not a cumulative distribution function. The first-order Edgeworth expansion is only accurate over the central domain.

4.5.4 Remarks

Accuracy of the Density Function Approximations

The saddle-point density function is non-negative, smooth and accurate in each example. The Edgeworth expansion density function is smooth and accurate around the mean, but is negative in the left tail. The saddle-point density function is better behaved and more accurate in the tails than the Edgeworth expansion density function. Both density functions are accurate around the mean.

The saddle-point density function is renormalised which makes it a probability density function. This is possible because the saddle-point density function is non-negative and decreases exponentially to 0 in the tails. The Edgeworth expansion density function cannot be renormalised without introducing severe bias due to its tendency to be negative in either tail.

Accuracy of the Distribution Function Approximations

The saddle-point distribution function has a range between 0 and 1 and is smooth and accurate in each example. The Edgeworth expansion is smooth and accurate around the mean, but is

negative in the left tail. The Edgeworth expansion and saddle-point distribution function are very similar around the mean.

Another approximation of the distribution function can be determined by integrating the normalised saddle-point density function. The saddle-point density function is not the derivative of the saddle-point distribution function - they are derived using different saddle-point methods as is discussed in sections 3.2 and 3.3. The distribution function corresponding to the normalised saddle-point density function would be smooth, non-negative and accurate but would not be of a closed form. Numerical methods are necessary to integrate such a function.

The Actual Distribution of $\tilde{H}_{k,n}$

The actual density and distribution functions of $\tilde{H}_{10,100}$ are unstable because $b_{100,10}$ is close to 0 in both examples. The size of $b_{100,10}$ implies that the denominator $\prod_{i \neq j}(\mu_j - \mu_i)$ is smaller than $b_{100,10}^{-10}$, which causes the density and distribution functions to "blow up".

4.6 Simulation Study

4.6.1 Introduction

The aim of this section is to investigate the behaviour of approximations of the density and distribution functions of the Hill estimator when the parameters $\gamma, b_{n,k} > 0$ and $\rho < 0$ are estimated. The approximations discussed in section 4.5 are considered, namely the Edgeworth expansion of Cuntz, Haeusler & Segers (2003) and the saddle-point density and distribution functions given by equation (4.13) and (4.14). These approximations are illustrated for a variety of heavy-tailed distributions, in each case where the parameters $\gamma > 0, b_{n,k} > 0$ and $\rho < 0$ are estimated.

This section has four parts. In the first part the parameter estimation methods are discussed. There are two such methods - the first is the method of maximum likelihood and the second is a least-squares regression fit. In the second part the estimated saddle-point and Edgeworth expansion density and distribution functions are defined. In the third part samples are simulated from given distributions and for each sample the parameters $\gamma, b_{n,k}$ and ρ are estimated. These parameter estimates are used to determine the estimated saddle-point and Edgeworth expansion density and distribution functions of the Hill estimator. In the fourth part the behaviour of the estimated saddle-point and Edgeworth expansion density and distribution functions are discussed.

4.6.2 Parameter Estimation

The parameters $\gamma, b_{n,k}$ and ρ are estimated using the second-order approximation (4.8):

$$Z_j \overset{D}{\approx} \left(\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}\right) E_j \qquad \text{for} \qquad 1, 2, \ldots, k$$

where k is small, n is large, $\gamma > 0, b_{n,k} > 0$ and $\rho < 0$. Two parameter estimation methods are discussed below.

Maximum Likelihood Estimation

Equation (4.8) implies that the density function of Z_j can be approximated by

$$f_{Z_j}(z) = \frac{\exp\left\{-\frac{z}{\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}}\right\}}{\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}} \qquad \text{for } z > 0$$

It follows that the log-likelihood of the sample $Z_1, Z_2, \ldots, Z_k$ can be approximated by

$$\sum_{j=1}^{k} \log f_{Z_j}(Z_j) \approx -\sum_{j=1}^{k} \log\left(\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}\right) - \sum_{j=1}^{k} \frac{z}{\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}} = \log L(\gamma, b_{n,k}, \rho)$$

The maximum likelihood estimators $\hat{\gamma}, \hat{b}_{n,k}$ and $\hat{\rho}$ are defined as

$$\left(\hat{\gamma}, \hat{b}_{n,k}, \hat{\rho}\right) = \arg \inf_{(\gamma, b_{n,k}, \rho)} \log L\left(\gamma, b_{n,k}, \rho\right) \qquad \text{such that} \qquad \hat{\gamma} > 0, \hat{b}_{n,k} > 0, \hat{\rho} < 0 \qquad (4.14)$$

The maximum likelihood estimators given by equation (4.15) cannot be solved explicitly by differentiating with respect to each parameter and setting each derivative equal to zero. Instead, equation (4.15) can be solved for a vector of observed Z_j's using the optim funtion in the programming language **R**.

Regression Estimation

Equation (4.8) can be written as

$$Z_j = \gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho} + \epsilon_j \qquad \text{for} \qquad j = 1, 2, \ldots, k$$

where $\epsilon_j = \left(\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}\right)(E_j - 1)$ is the error term with mean 0 and variance $\left(\gamma + b_{n,k}\left(\frac{j}{k}\right)^{-\rho}\right)^2$. Note that the variance of the error term ϵ_j is maximised when $\rho = 0$.

We wish to fit a least squares line where $\{Z_j\}$ is the response and $\left\{\left(\frac{j}{k}\right)^{-\rho}\right\}$ is the covariate, and consequently, γ is the intercept and $b_{n,k}$ is the gradient. In order to have an optimal fit, the parameter $\rho < 0$ is chosen to maximise the positive linear relationship between the response $\{Z_j\}$ and the covariate $\left\{\left(\frac{j}{k}\right)^{-\rho}\right\}$ as follows:

$$\hat{\rho} = \arg\sup_{\rho} \max\left\{Corr\left(\{Z_j\}, \left\{\left(\frac{j}{k}\right)^{-\rho}\right\}\right), 0\right\} \tag{4.15}$$

Equation (4.16) can be solved for a vector of observed Z_j's using the optim funtion in the programming language **R**. The parameters γ and $b_{n,k}$ are estimated as the intercept and gradient of the linear least-squares fit of $\{Z_j\}$ on $\left\{\left(\frac{j}{k}\right)^{-\hat{\rho}}\right\}$ respectively.

4.6.3 The Estimated Approximations of the Density and Distribution Functions of the Hill Estimator

Both the saddle-point and the Edgeworth expansion density and distribution functions have parameters $\gamma, b_{n,k}$ and ρ. If the distribution function $F(x)$ of the random sample $X_1, X_2, \ldots, X_n$ is known and of a closed form, then the parameters $\gamma, b_{n,k}$ and ρ can be determined explicitly as is the case in section 4.5. However, if the distribution function $F(x)$ is unknown, then it is necessary to estimate the parameters $\gamma, b_{n,k}$ and ρ using one of the parameter estimation methods discussed in section 4.6.2. These estimation methods assume that $F(x)$ is Pareto-type (in other words $F(x)$ has a heavy tail) and that the slowly-varying part of the tail quantile function satisfies $C_\rho(b)$ for some $\rho < 0$ where $b(x)$ is regularly varying with index ρ.

The estimated saddle-point and Edgeworth expansion density and distribution functions are determined by substituting the parameter estimates $\hat{\gamma}, \hat{b}_{n,k}$ and $\hat{\rho}$ into the formulae for these functions which are given by equations (4.10), (4.9), (4.13) and (4.14) respectively. These functions are estimated for a range of samples from a known distribution. The aim of estimating these functions is to investigate their behaviour for estimated parameter values that differ substantially from the actual parameter values.

4.6.4 The Burr Case

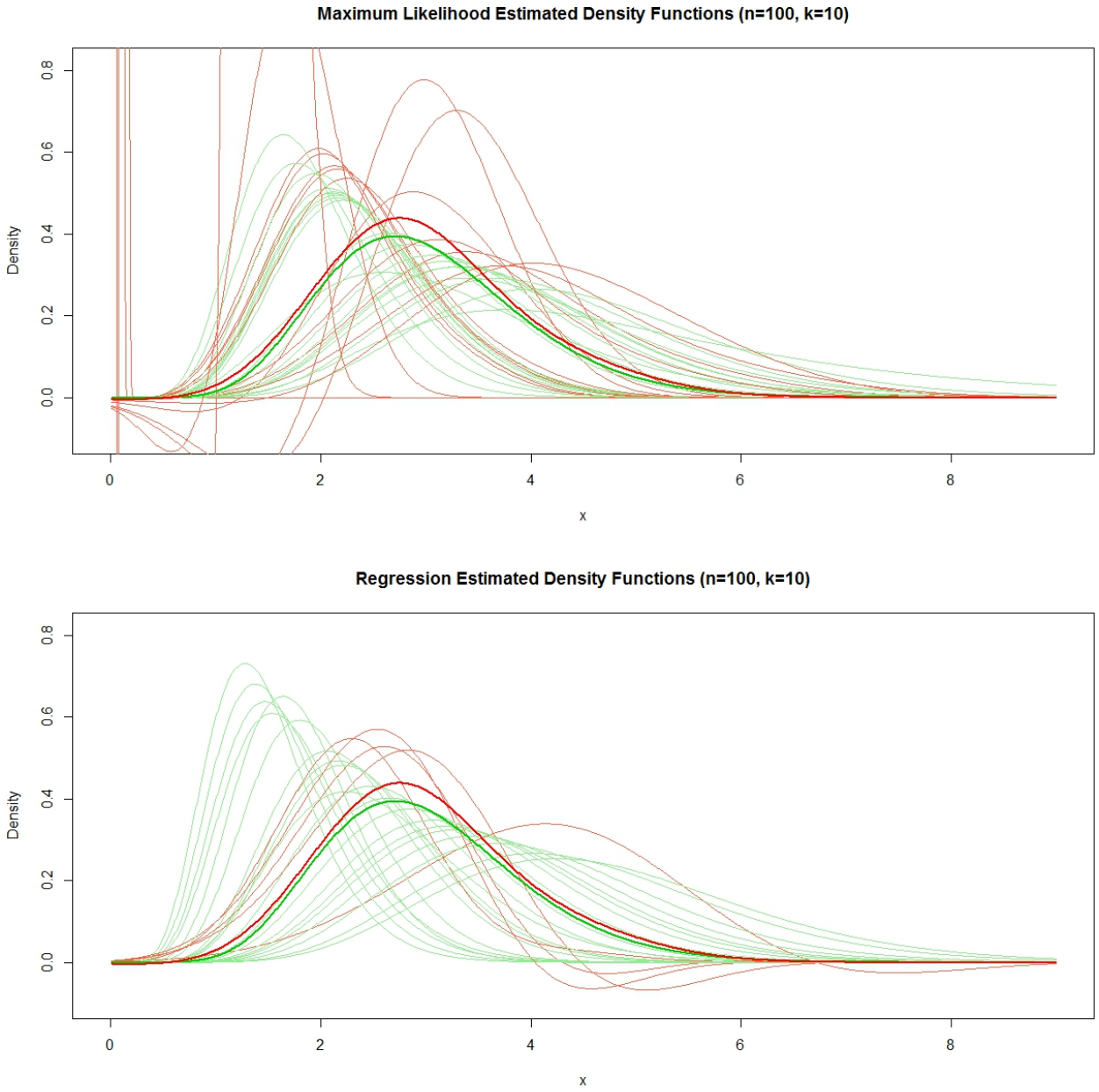


Figure 4.5: The 20 light green and 20 light red lines represent the estimated saddle-point and Edgeworth expansion density functions respectively. Each of these 20 sets of lines correspond to a generated $Burr\left(\frac{1}{2}, \frac{2}{3}\right)$ sample. The thicker green and red lines represent the actual saddle-point and Edgeworth expansion density functions respectively.

The Edgeworth density functions are unstable for the maximum likelihood parameter estimates and some take on negative values in the left tail. This is partly due to the inaccurate and unstable parameter estimates which the maximum likelihood method tends to yield. The Edgeworth density functions are stable for the regression parameter estimates but some take on negative values in the right tail. Figure 4.7 illustrates how the Edgeworth expansion density functions can take on negative values in either tail depending on the parameter values.

All the saddle-point density functions are non-negative and decrease to 0 in the tails. The saddle-point density functions are therefore well-behaved for both the maximum likelihood

estimates and regression estimates.

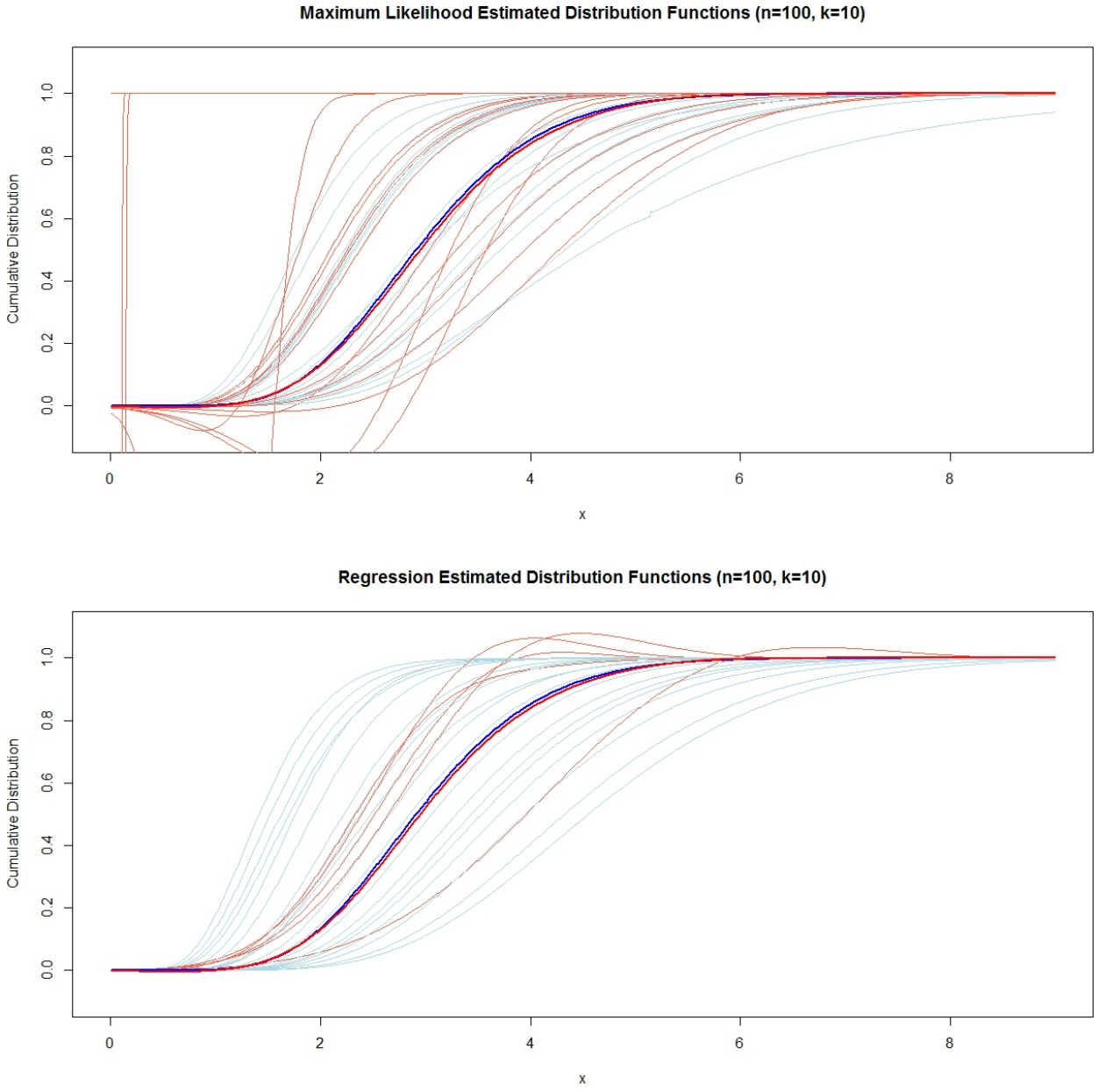


Figure 4.6: The 20 light blue and 20 light red lines represent the estimated saddle-point and Edgeworth expansion distribution functions respectively. Each of these 20 sets of lines correspond with a generated $Burr\left(\frac{1}{2}, \frac{2}{3}\right)$ sample. The thicker blue and red lines represent the actual saddle-point and Edgeworth expansion distribution functions respectively.

The Edgeworth distribution functions mimic the behaviour of their corresponding density functions. They are highly unstable or even degenerate for the maximum likelihood estimates and take on negative values in left the tail. For the regression estimates, the Edgeworth distribution function takes on values larger than 1. The well-behaved Edgeworth expansion distribution functions are accurate around the mean.

The saddle-point distribution functions are continuous, smooth and well-behaved, each with a range of $[0, 1]$. Furthermore, the saddle-point distribution functions are very similar to their Edgeworth expansion counterparts around the mean.

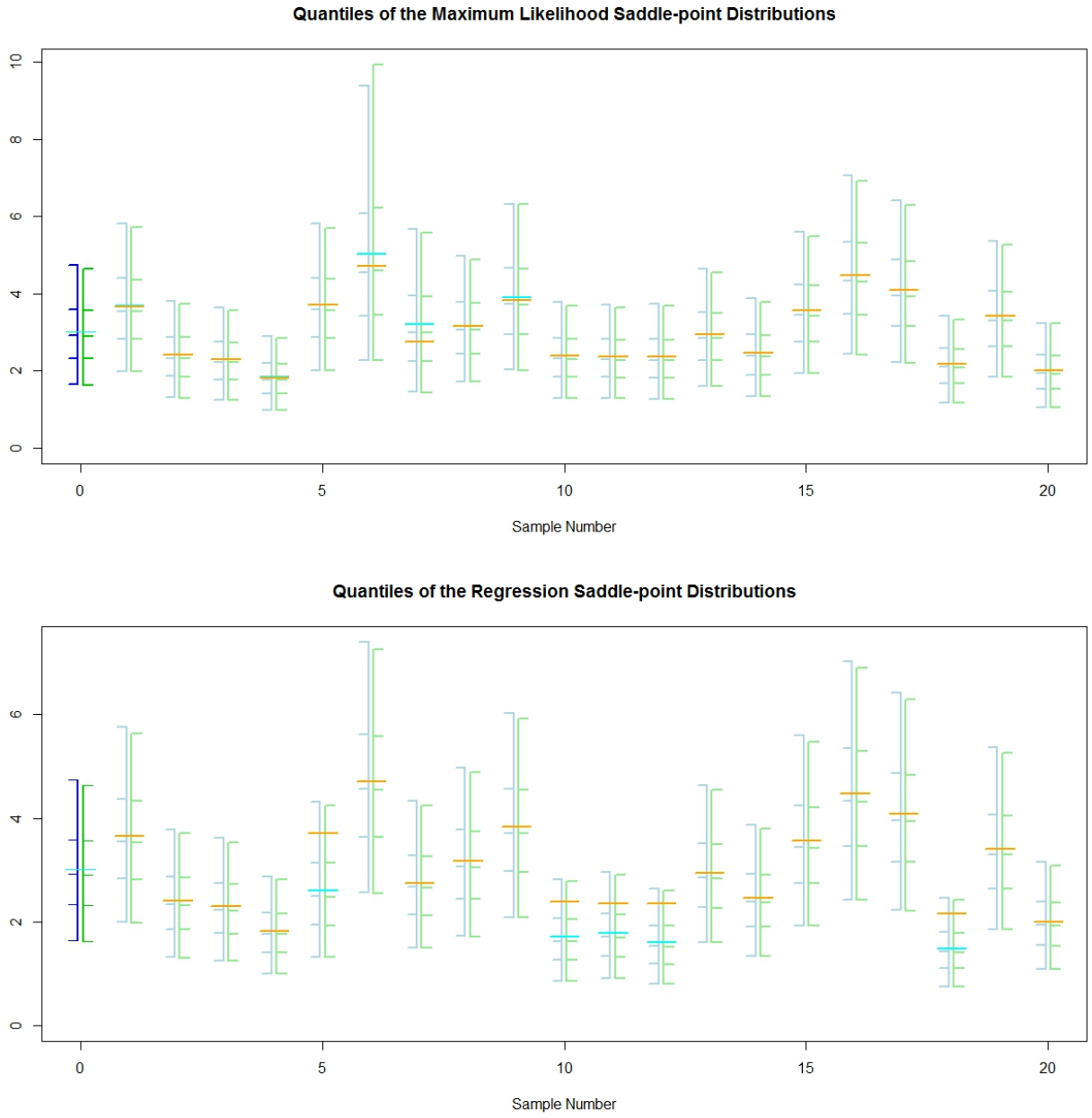


Figure 4.7: The graph illustrates the .05, .25, .5, .75 and .95 quantiles of the saddle-point density and distribution functions. The dark green and blue lines represent the quantiles of the actual saddle-point density and distribution functions respectively. The light green and blue lines represent the quantiles of the estimated saddle-point density and distribution functions respectively, for each sample generated. The cyan line represents the saddle-point mean and the orange line represents the Hill estimate for each sample.

The boxplot of the quantiles of the actual and estimated saddle-point density and distribution functions illustrate the similarities and differences between them. The above graph yields the following insights:

• The saddle-point density and distribution functions appear to be very similar, since for each sample their quantiles are very close to one another.

• The estimated saddle-point density and distribution functions vary significantly between the samples. This is partly due to sampling error, since a small sample of size 10 is chosen.

• The Hill estimator is in the center of the maximum likelihood estimated saddle-point distributions, but the same is not true for the regression estimated saddle-point distributions. This

suggests that the maximum likelihood estimates yield more accurate saddle-point density and distribution functions.

4.6.5 The Frechet Case

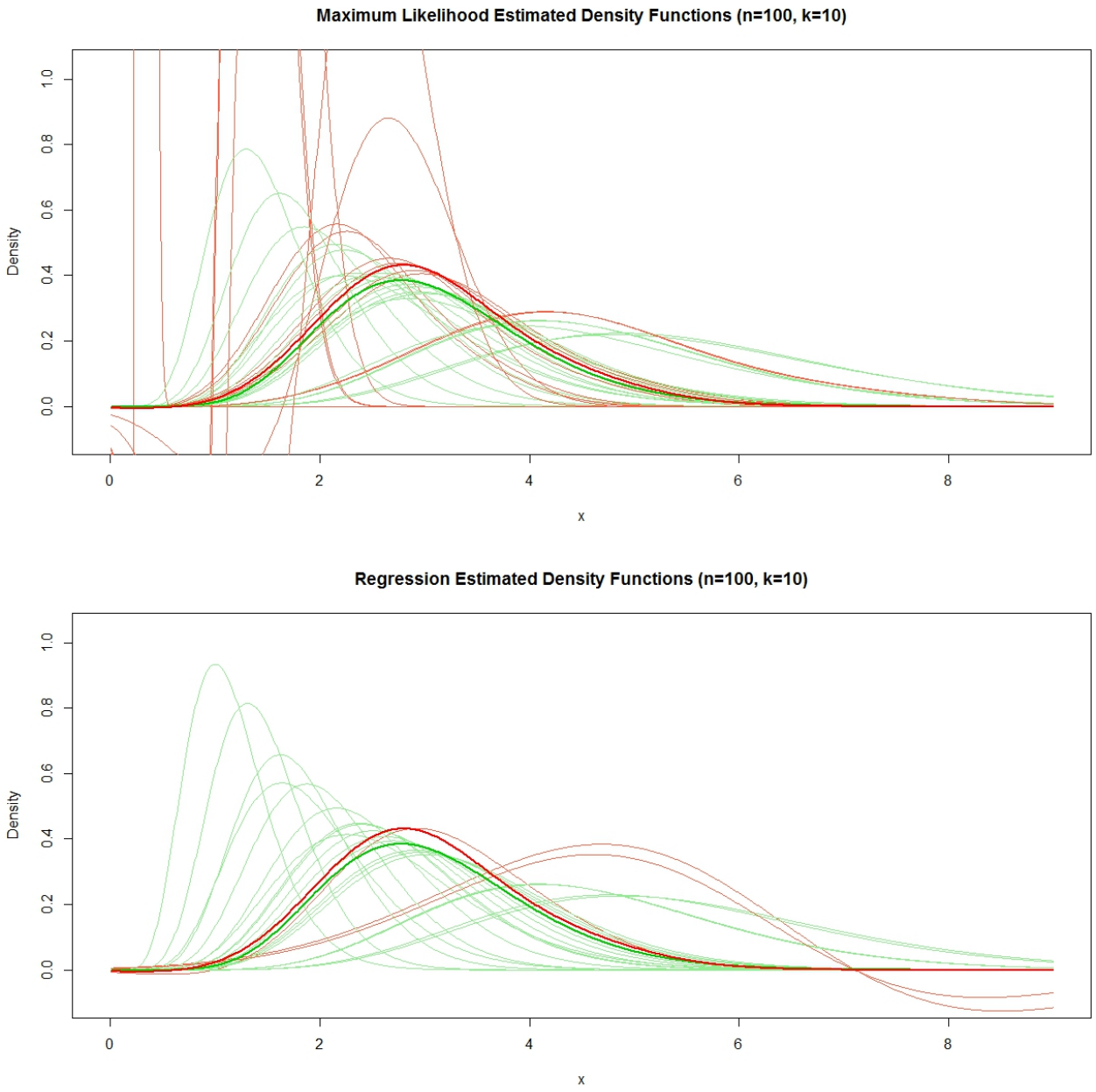


Figure 4.8: The 20 light green and 20 light red lines represent the estimated saddle-point and Edgeworth expansion density functions respectively. Each of these 20 sets of lines correspond with a generated $Frechet\left(\frac{1}{3}\right)$ sample. The thicker green and red lines represent the actual saddle-point and Edgeworth expansion density functions respectively.

Again the Edgeworth density functions are highly unstable for the maximum likelihood parameter estimates and some take on negative values in the left tail. The Edgeworth density functions are stable for the regression parameter estimates but some take on negative values in the right tail.

The saddle-point density functions are all non-negative and decrease to 0 in the tails. The saddle-point density functions are therefore well-behaved for both the maximum likelihood and regression estimates.

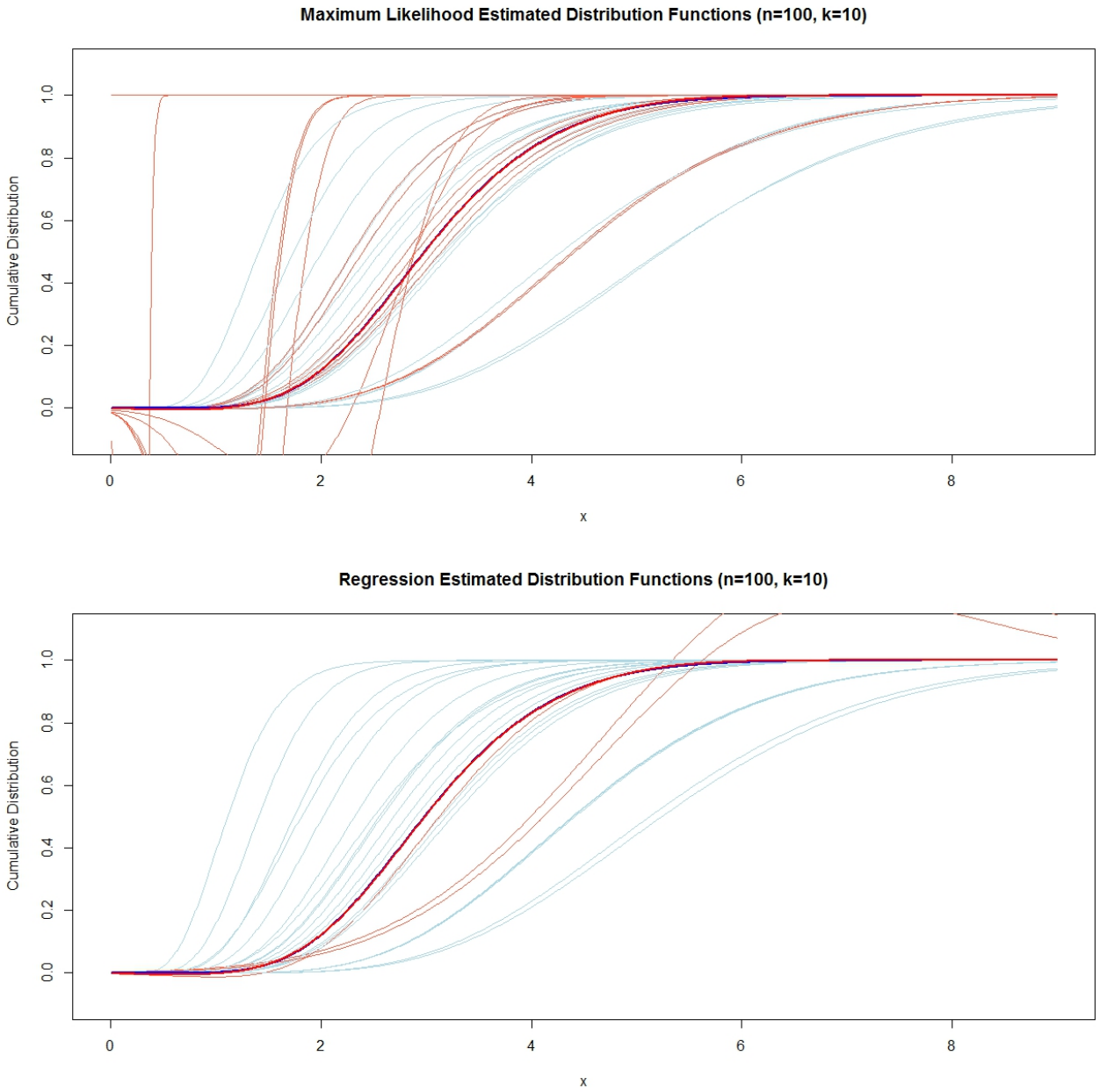


Figure 4.9: The 20 light blue and 20 light red lines represent the estimated saddle-point and Edgeworth expansion distribution functions respectively. Each of these 20 sets of lines correspond with a generated $Frechet\left(\frac{1}{3}\right)$ sample. The thicker blue and red lines represent the actual saddle-point and Edgeworth expansion distribution functions respectively.

Again the Edgeworth distribution functions mimic the behaviour of their corresponding density functions. They are highly unstable or even degenerate for the maximum likelihood estimates and take on negative values in left the tail. For the regression estimates, the Edgeworth distribution function takes on values larger than 1. The well-behaved Edgeworth expansion distribution functions are accurate around the mean.

The saddle-point distribution functions are continuous, smooth and well-behaved, each with a range of $[0, 1]$.

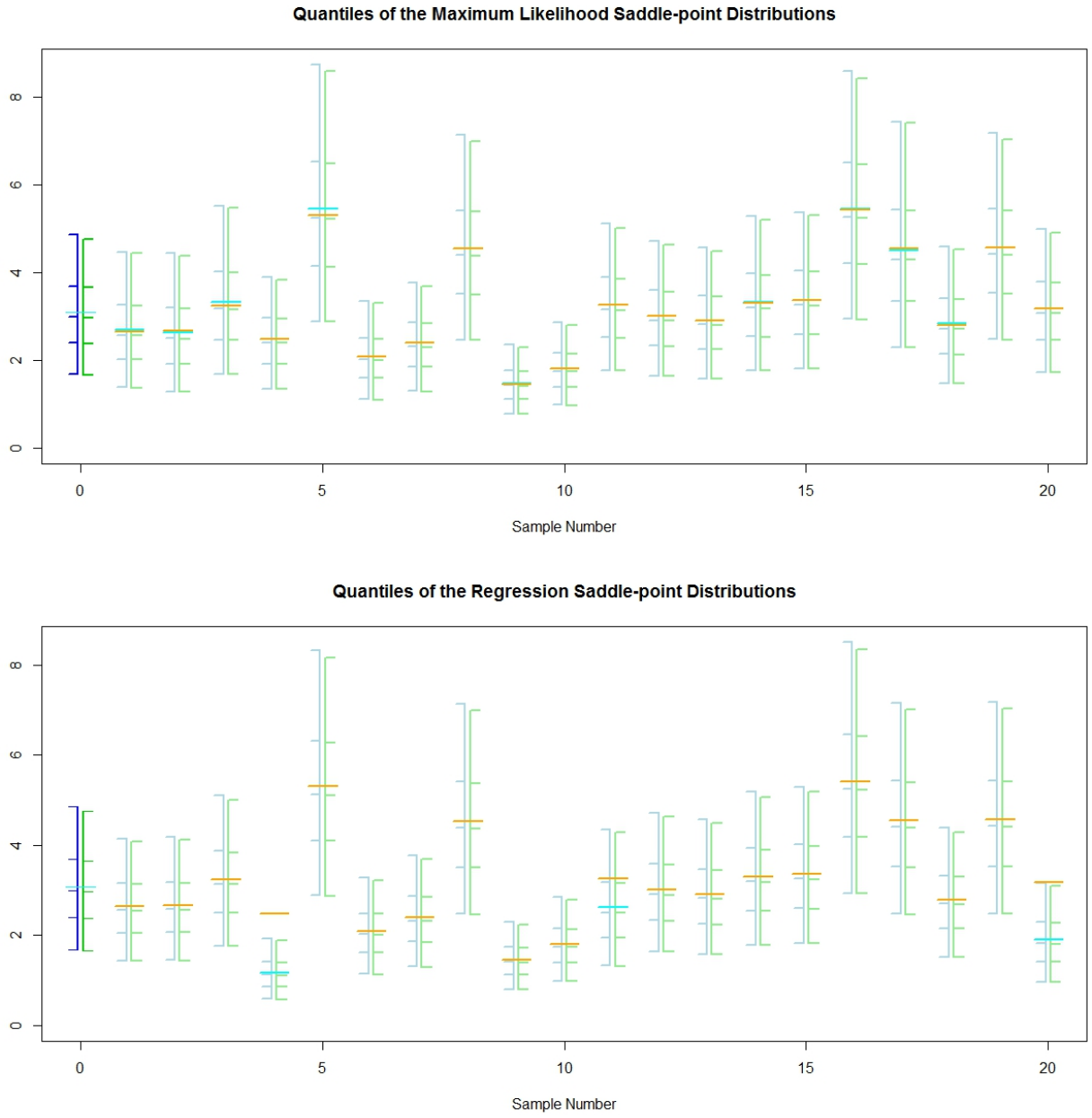


Figure 4.10: The graph illustrates the .05, .25, .5, .75 and .95 quantiles of the saddle-point density and distribution functions. The dark green and blue lines represent the quantiles of the actual saddle-point density and distribution functions respectively. The light green and blue lines represent the quantiles of the estimated saddle-point density and distribution functions respectively, for each sample generated. The cyan line represents the saddle-point mean and the orange line represents the Hill estimate for each sample.

The above graph yields the following insights, which are similar to the Burr case:
• The saddle-point density and distribution functions appear to be very similar, since for each sample their quantiles are very close to one another.
• The estimated saddle-point density and distribution functions vary significantly between the samples. This is partly due to sampling error, since a small sample of size 10 is chosen.
• The Hill estimator is in the center of the maximum likelihood estimated saddle-point distributions, but the same is not true for the regression estimated saddle-point distributions. This suggests that the maximum likelihood estimates yield more accurate saddle-point density and distribution functions.

4.6.6 The Inverse Gamma Case

The inverse gamma distribution does not have a closed-form distribution function and it is therefore not possible to determine the actual values of the parameters $\gamma, b_{n,k}$ and ρ explicitly. The parameters $\gamma, b_{n,k}$ and ρ can be estimated using the two estimation methods of section 4.6.2 because the inverse gamma distribution is of Pareto-type and satisfies the second-order conditions of section 4.4.2.

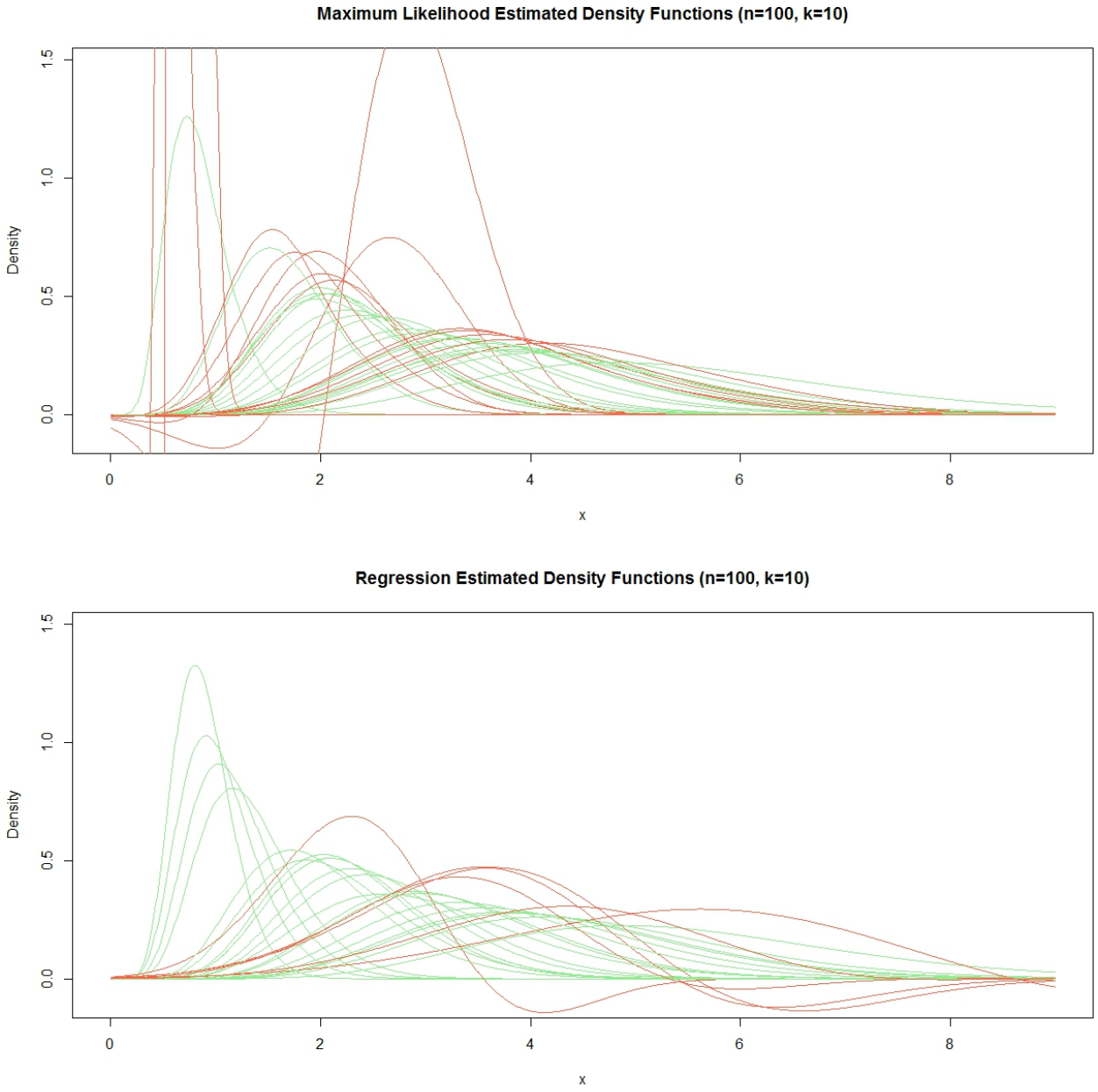


Figure 4.11: The 20 light green and 20 light red lines represent the estimated saddle-point and Edgeworth expansion density functions respectively. Each of these 20 sets of lines correspond with a generated $Inv\Gamma\left(1, \frac{1}{3}\right)$ sample.

The Edgeworth expansion density functions are again stable for the regression parameters but highly unstable for the maximum likelihood parameter estimates. The Edgeworth density functions are also negative in the left and right tails for the maximum likelihood and regression parameter estimates respectively.

The saddle-point density functions are all non-negative and decrease to 0 exponentially in the tails, and are therefore well-behaved for both the maximum likelihood estimates and the regression estimates.

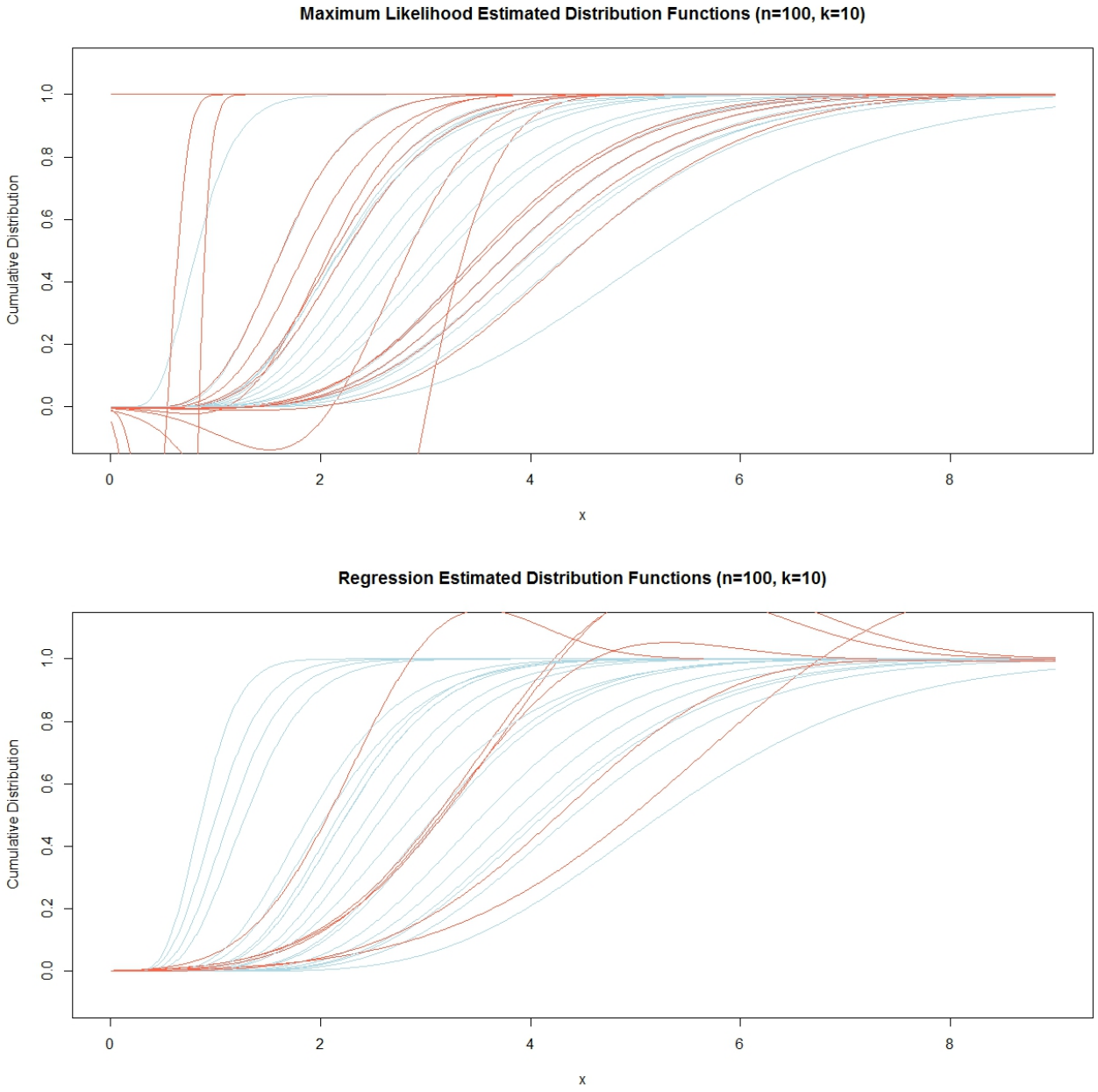


Figure 4.12: The 20 light blue and 20 light red lines represent the estimated saddle-point and Edgeworth expansion distribution functions respectively. Each of these 20 sets of lines correspond with a generated $Inv\Gamma\left(1, \frac{1}{3}\right)$ sample.

Again the Edgeworth distribution functions are highly unstable or degenerate for the maximum likelihood estimates and take on negative values in left the tail. For the regression estimates, the Edgeworth distribution function takes on values larger than 1.

The saddle-point distribution functions are continuous, smooth and well-behaved, each with a range of $[0, 1]$.

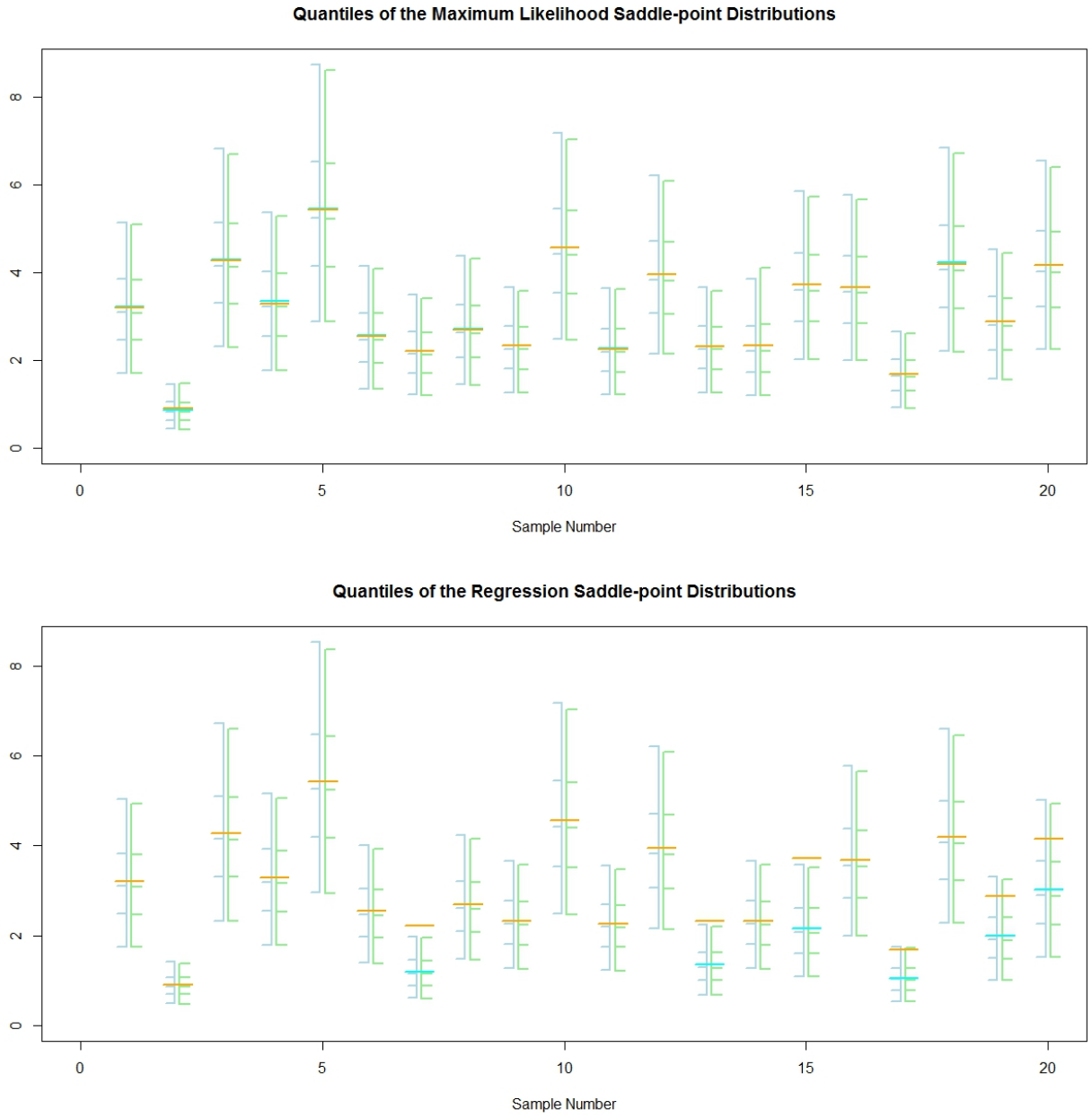


Figure 4.13: The graph illustrates the .05, .25, .5, .75 and .95 quantiles of the saddle-point density and distribution functions. The light green and blue lines represent the quantiles of the estimated saddle-point density and distribution functions respectively, for each sample generated. The cyan line represents the saddle-point mean and the orange line represents the Hill estimate for each sample.

The above graph yields the following insights, which are similar to both the Burr and Frechet cases:

• The saddle-point density and distribution functions appear to be very similar, since for each sample their quantiles are very close to one another.

• The estimated saddle-point density and distribution functions vary significantly between the samples. This is partly due to sampling error, since a small sample of size 10 is chosen.

• The Hill estimator is in the center of the maximum likelihood estimated saddle-point distributions, but the same is not true for the regression estimated saddle-point distributions. This suggests that the maximum likelihood estimates yield more accurate saddle-point density and distribution functions.

4.6.7 The Log Gamma Case

The log-gamma case is similar to the inverse gamma case. Both distributions do not have closed-form distribution functions and therefore the actual values of the parameters $\gamma, b_{n,k}$ and ρ cannot be determined explicitly.

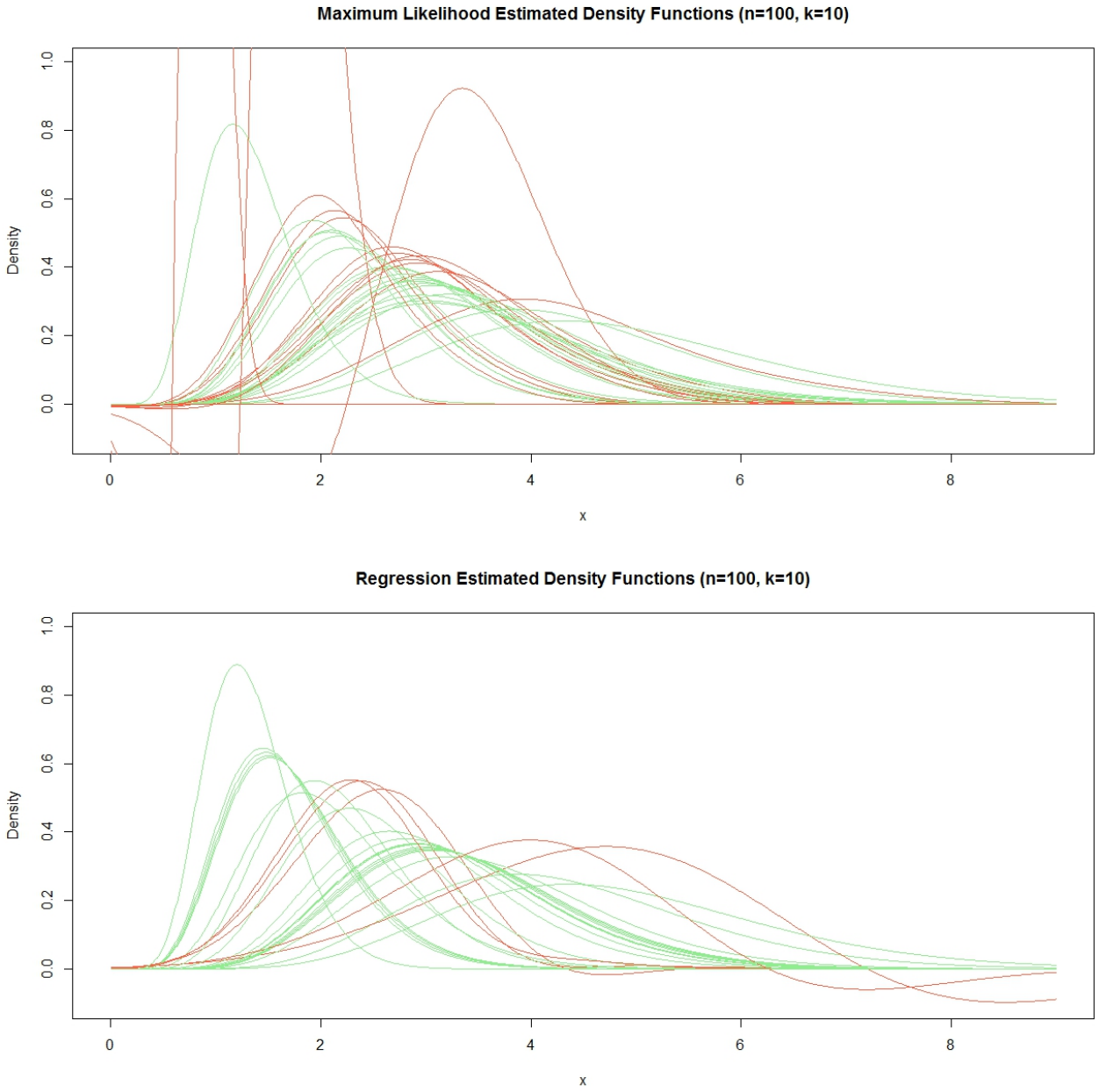


Figure 4.14: The 20 light green and 20 light red lines represent the estimated saddle-point and Edgeworth expansion density functions respectively. Each of these 20 sets of lines correspond with a generated $\log \Gamma\left(\frac{1}{3}, 1\right)$ sample.

The Edgeworth expansion density functions are again stable for the regression parameters but highly unstable for the maximum likelihood parameter estimates. The Edgeworth density functions are also negative in the left and right tails for the maximum likelihood and regression parameter estimates respectively.

The saddle-point density functions are all non-negative and decrease to 0 exponentially in the tails, and are therefore well-behaved for both the maximum likelihood estimates and the

regression estimates.

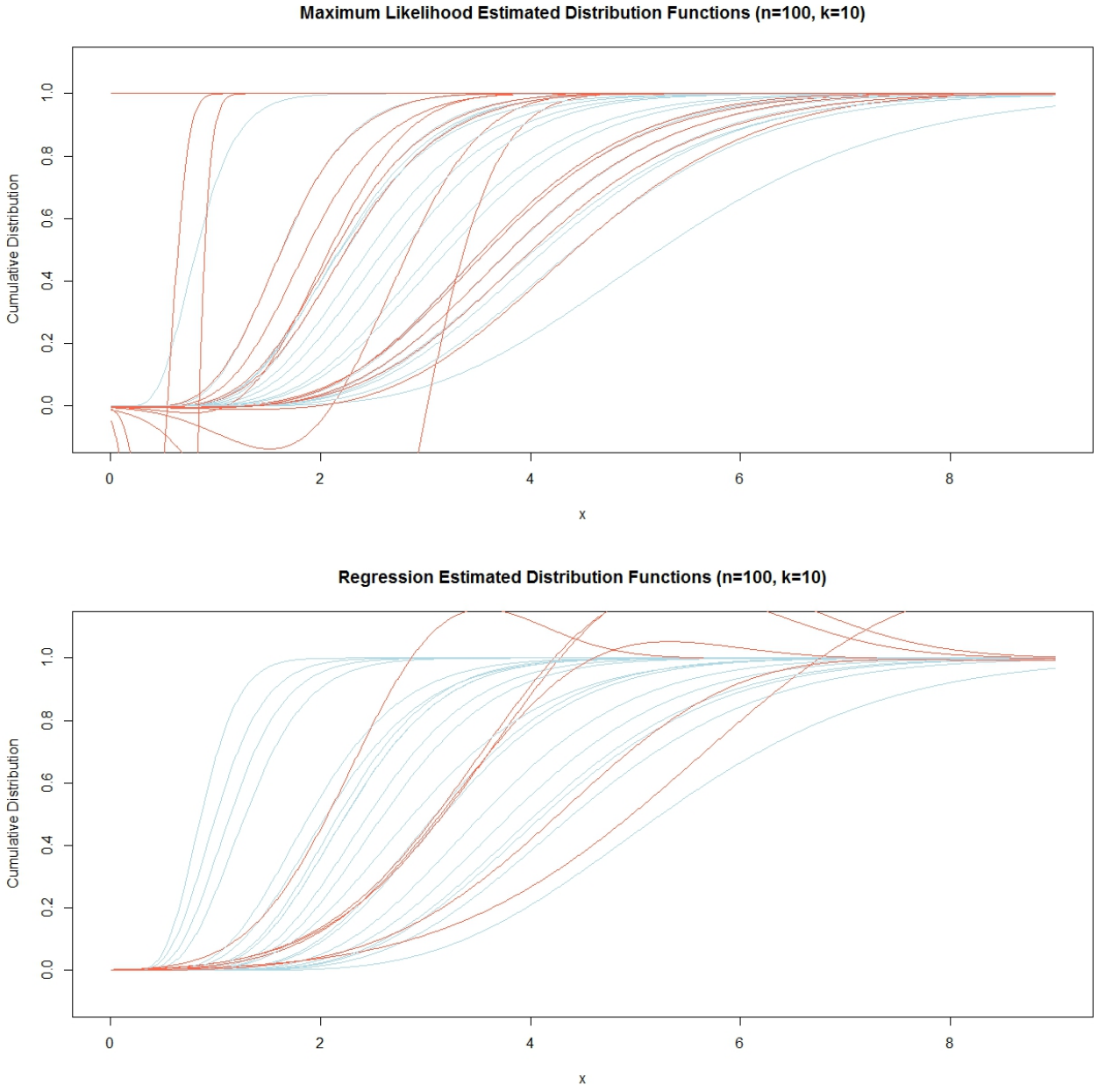


Figure 4.15: The light blue and red lines represent the estimated saddle-point and Edgeworth expansion distribution functions respectively. Each of these 20 sets of lines correspond with a generated $\log \Gamma\left(\frac{1}{3}, 1\right)$ sample.

Again the Edgeworth distribution functions are highly unstable or degenerate for the maximum likelihood estimates and take on negative values in left the tail. For the regression estimates, the Edgeworth distribution function takes on values larger than 1.

The saddle-point distribution functions are continuous, smooth and well-behaved, each with a range of $[0, 1]$.

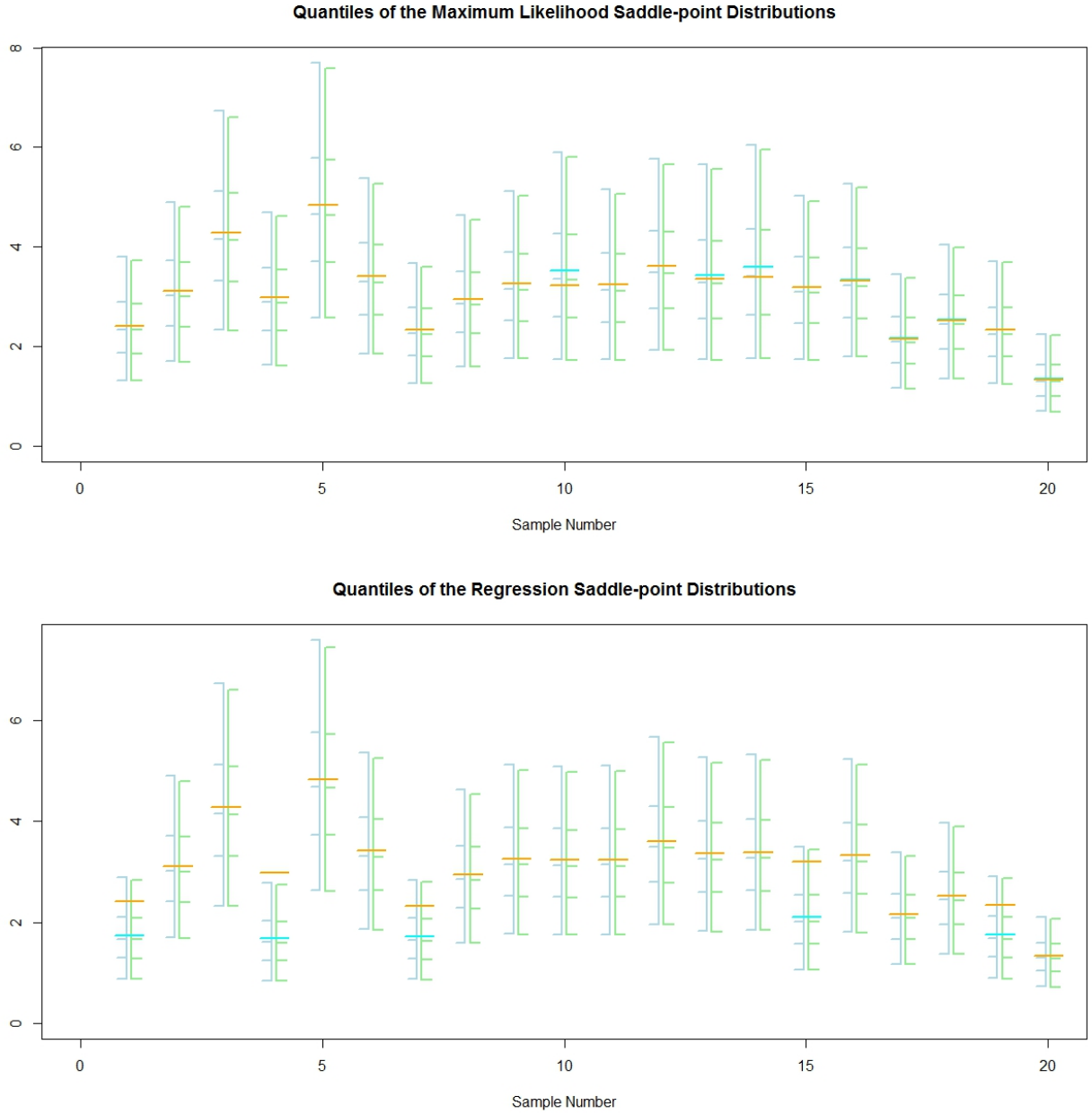


Figure 4.16: The graph illustrates the .05, .25, .5, .75 and .95 quantiles of the saddle-point density and distribution functions. The light green and blue lines represent the quantiles of the estimated saddle-point density and distribution functions respectively, for each sample generated. The cyan line represents the saddle-point mean and the orange line represents the Hill estimate for each sample.

The above graph yields the following insights, which are similar to the previous distributions:
• The saddle-point density and distribution functions appear to be very similar, since for each sample their quantiles are very close to one another.
• The estimated saddle-point density and distribution functions vary significantly between the samples. This is partly due to sampling error, since a small sample of size 10 is chosen.
• The Hill estimator is in the center of the maximum likelihood estimated saddle-point distributions, but the same is not true for the regression estimated saddle-point distributions. This suggests that the maximum likelihood estimates yield more accurate saddle-point density and distribution functions.

4.7 Summary

4.7.1 Introduction

This summative section discusses the behaviour and accuracy of the saddle-point and Edgeworth expansion density and distribution functions which are investigated in sections 4.5 and 4.6.

4.7.2 Behaviour

The Saddle-point Density and Distribution Functions

Both the saddle-point density and distribution functions are well-behaved, even in scenarios when the parameters $\gamma, b_{n,k}$ and ρ are estimated. The saddle-point density function is non-negative and smooth, and decreases to 0 in the tails. Due to its good behaviour, the saddle-point density function can be renormalised so that its integral equals exactly 1. The saddle-point distribution function is smooth, monotone-increasing and has the range $[0, 1]$. The saddle-point distribution function is smooth when the actual parameter values for $\gamma, b_{n,k}$ and ρ are used. However, when the values of $\gamma, b_{n,k}$ and ρ are estimated, then the corresponding values of the saddle-point $\hat{t}$ close to 0 can cause the term $\frac{1}{\hat{t}\sqrt{K^{(2)}(\hat{t})}} - \frac{1}{\hat{w}}$ in equation (4.14) to be discontinuous around the mean. This is partly due to the fact that the saddle-point $\hat{t}$ is calculated numerically for given values of $\gamma, b_{n,k}$ and ρ, and that the estimated values of $\gamma, b_{n,k}$ and ρ are also calculated numerically (Butler, 2007:12). Numerical calculations introduce rounding errors. The discontinuities of the saddle-point distribution function are removed by choosing a window around the mean over which the distribution function is made linear and continuous.

The Edgeworth Expansion

The behaviour of the Edgeworth expansion distribution function is dependent on the parameter values for $\gamma, b_{n,k}$ and ρ:

- For the actual parameter values, the Edgeworth expansion distribution function is well-behaved around the mean but is problematic in the tails. In many scenarios the Edgeworth distribution function is negative in the left tail and/or larger than 1 in the right tail.
- For maximum likelihood parameter estimates, the Edgeworth expansion distribution function can be degenerate, possibly due to γ estimates close to 0. In the cases when the Edgeworth expansion distribution function is non-degenerate, it tends to be well-behaved around the mean and in the right tail but it is negative in the left tail.
- For regression parameter estimates, the Edgeworth expansion distribution is well-behaved around the mean and in the left tail but it is larger than 1 in the right tail. The regression parameter estimates of γ are more centered around the actual value of γ than the maximum likelihood estimates. This is one of the reasons why the Edgeworth expansion distribution function is better behaved for regression estimated parameters.

4.7.3 Accuracy

There are two levels of accuracy when it comes to evaluating an approximation of a density or distribution function.

The first level of accuracy measures how close the approximation is to the density or distribution function when the values of the parameters $\gamma, b_{n,k}$ and ρ are known. This level of

accuracy is illustrated in section 4.5 where the density and distribution function of the Hill estimator are plotted against the saddle-point and Edgeworth density and distribution functions respectively. It is concluded from the results in section 4.5 that the actual saddle-point density and distribution functions are more accurate and well-behaved than the actual Edgeworth expansion density and distribution functions.

The second level of accuracy measures how close the actual approximation is to the estimated approximation. The values of the parameters $\gamma, b_{n,k}$ and ρ are assumed to be unknown for the estimated approximation and they therefore have to be estimated based on a sample. This is illustrated in section 4.6 where the actual saddle-point and Edgeworth density and distribution functions are plotted against their estimated counterparts.

Two conclusions can be made from section 4.6. Firstly, the estimated saddle-point density and distributions functions are far more accurate and well-behaved than the estimated Edgeworth expansion density and distribution functions. Secondly, the method of maximum likelihood yields more accurate saddle-point density and distribution function estimates than does the regression estimation method. The maximum likelihood estimated saddle-point density and distribution functions, however, differ significantly from the actual saddle-point density and distribution functions This is mostly due to sampling error, since a small sample of size of 10 is chosen to estimate the parameters.

Chapter 5

Conclusion

5.1 Introduction

This chapter serves as a summary of all the key findings of the thesis and is structured around four central topics.

The first topic concerns the utility of the saddle-point method. The saddle-point method has a wide variety of applications in mathematics and applied mathematics. It can be used to yield asymptotic estimates of special functions such as Bessel functions, the hypergeometric function and Riemann's zeta function. Furthermore, the saddle-point method is an approach which can be extended and adapted to widen its applications. This is evident in the saddle-point method of de Bruijn (1958), Bleistein (1966) and Flajolet & Sedgewick (2009). The utility of the saddle-point method is discussed in section 5.2.

The second topic concerns the high accuracy and good behaviour of the saddle-point approximation in a non-asymptotic setting. This mystery is a central topic of this thesis and is illustrated by the examples in sections 3.2.6 and 3.3.5. The saddle-point density and distribution function of the Hill estimator, which are investigated in sections 4.5 and 4.6, also illustrate the accuracy and good behaviour of the saddle-point approximation for a small sample size of 10. The saddle-point approximation in a non-asymptotic setting is discussed in section 5.3.

The third topic concerns the saddle-point approximation of the distribution of the Hill estimator. The saddle-point density and distribution functions of the Hill estimator are proposed in section 4.4.6 and their accuracy and behaviour are investigated in sections 4.5 and 4.6. The saddle-point density and distribution functions of the Hill estimator are discussed in section 5.4.

The fourth topic concerns the two approaches to deriving the saddle-point approximation: the saddle-point method and the indirect Edgeworth expansion. Both of these approaches provide unique insights and are briefly discussed in section 5.5.

The final section highlights three areas of further research regarding the saddle-point approximation in the field of statistics.

5.2 The Utility of the Saddle-point Method

Before discussing the merits of the saddle-point method applied to the statistical context (yielding the saddle-point approximation), much insight is to be gained by appreciating the variety of ways in which the saddle-point method is applied.

5.2.1 Applications in a Variety of Fields

The saddle-point method is very useful in providing asymptotic estimates of solutions to differential equations. Such solutions include the hypergeometric function and the Bessel functions, which include the Hankel functions and the Airy functions. The saddle-point method is also very useful in providing asymptotic estimates of Cauchy coefficient integrals and the Riemann-Siegel function, which is related to the Riemann zeta function.

The saddle-point method can be applied to the Fourier inversion formula to yield a highly accurate approximation. This is especially useful in cases where the inverse Fourier transform is difficult to calculate analytically.

The saddle-point method has such a wide range of applications because of two reasons: the solutions to many real problems are more easily found using complex methods, and the saddle-point method can be extended to include a larger variety of complex integrals.

5.2.2 Using Complex Methods to Solve Real Problems

Methods in complex analysis are often very useful in solving real problems. Complex integration can be used as a means to solve real integrals and infinite sums. For example, the value of the infinite sum $\sum_{n=1}^{\infty} \frac{1}{n^2}$ is found to equal $\frac{\pi^2}{6}$ by using the Residue Theorem (Conway, 1973).

Fourier analysis is central to applied mathematics, physics and statistics. By applying the Fourier transform to a variable observed over time, the variable is transformed to the frequency domain, and many questions regarding the variable are more easily answered by observing it in the frequency domain. For example, a signal is a variable which can be observed over time. By applying the Fourier transform to a signal it is transformed to the frequency domain, in which it is apparent which frequencies the signal takes on. A filter can be applied to eliminate unwanted frequencies in the frequency domain, and the Fourier inversion formula can be applied to transform this filtered signal back into the time domain. Both the Fourier transformation and the inverse Fourier transformation can be written as complex integrals.

Complex integrals and their properties prove to be useful in solving real problems. This implies that the saddle-point method, which is an invaluable tool in approximating complex integrals, can be applied to solve real problems.

5.2.3 Extensions of the Saddle-point Method

The saddle-point method is an approach to estimating a complex integral by choosing a path which traverses one or many saddle-points and then estimating the integrand locally near this or these saddle-points (Flajolet & Sedgewick, 2009:541). The saddle-point method is historically seen as an extension of Laplace's method to the complex space. A derivation of this is given by de Bruijn (1958) who uses a path of steepest descent to derive an asymptotic expansion of a complex integral which has a Laplacian form (see section 2.3.5). Bleistein (1966) extends the saddle-point method to yield an asymptotic expansion of a complex integral which is Laplacian

but with a pole at the origin.

Flajolet & Sedgewick (2009) devise a saddle-point method which is applicable in a more general setting than that of de Bruijn (1958). If a complex integral satisfies three weak conditions, then the saddle-point method yields an asymptotic estimate of the complex integral. Although this saddle-point method holds more generally, the result is simply an asymptotic estimate and not an asymptotic expansion as is the case of de Bruijn (1958).

5.3 The Saddle-point Approximation in a Non-asymptotic Setting

The saddle-point method is an asymptotic method and yields an asymptotic result. The saddle-point approximation is historically regarded as an asymptotic estimate (Daniels (1954), Barndorff-Nielsen & Cox (1979), Lugannani & Rice (1980)) because it is the result of applying the saddle-point method on the Fourier inversion formula where the cumulant generating function depended on a large sample size. The saddle-point approximation, however, is a very accurate approximation of the distribution of a random variable which does not depend on any large parameter, including sample size (Hougaard (1988), Butler (2007)).

5.3.1 The Accuracy of the Saddle-point Approximation in the Non-asymptotic Setting

Firstly, it is necessary to discuss the accuracy of the saddle-point approximation in a non-asymptotic setting. Hougaard (1988) is the first to comment on the accuracy of the saddle-point approximation in such a setting. In a comment on the paper by Reid (1988), he uses the saddle-point approximation to approximate the density and distribution function of the non-central chi-squared random variable. The saddle-point approximation is so accurate in this and many other cases that Hougaard (1988:231) concludes that "if the saddlepoint approximation is simpler than the exact distribution, we rarely need the exact distribution".

Butler (2007) also stresses the general setting in which the saddle-point approximation can be used. In his definition of the saddle-point density and distribution functions, he simply requires the cumulant generating function of the random variable to be defined on an interval containing 0. Butler (2007) further illustrates the accuracy of the saddle-point approximation by using non-asymptotic examples.

The saddle-point approximation of the density and distribution function of the Hill estimator, which are discussed in section 4.5, illustrates that the saddle-point approximation of the density and distribution functions are very accurate for a small sample size (of 10). This again emphasises the high accuracy of the saddle-point approximation in a non-asymptotic setting.

5.3.2 Lack of Theoretical Support

The saddle-point approximation is theoretically an asymptotic estimate and can therefore, according to theory, only be applied to random variables that depend on some large parameter, usually the sample size. In this asymptotic case, the saddle-point approximation proves to be highly accurate. However, the saddle-point approximation also proves to be highly accurate in the case where there is no large parameter, as is discussed in section 5.3.1, while there is little or no theory to support this. The high accuracy of the saddle-point approximation in the non-asymptotic setting is the prime mystery surrounding the method.

It is clear that the saddle-point approximation of the density and distribution functions depend solely on the quadratic approximation (3.2) and the linear approximation (3.7) respectively. If these approximations are asymptotic, then the corresponding saddle-point approximation of the density and distribution functions are asymptotic. If these approximations are non-asymptotic, then it is necessary to establish some measure of how good these approximations are, and, in turn, how accurate the corresponding saddle-point density and distribution functions are. However, due to the fact that the saddle-point density and distributions are

highly accurate in most cases, it is difficult to empirically establish both the measure of the adequacy of the approximations and the level of adequacy necessary to lead to accurate saddle-point approximations.

Rigorous theory needs to be developed to support the high accuracy of the saddle-point approximation in the non-asymptotic setting. Empirical study of the saddle-point approximation does not seem to glean the reasons for its accuracy - it just works. By returning to Butler (2007), the saddle-point approximation is perhaps best summarised by the opening line of his book: "Among the various tools that have been developed for use in statistics and probability over the years, perhaps the least understood and most remarkable tool is the saddlepoint approximation."

5.4 The Saddle-point Approximation of the Distribution of the Hill Estimator

The saddle-point density and distribution functions of the Hill estimator are proposed in section 4.4 and their accuracy and behaviour are illustrated in sections 4.5 and 4.6.

5.4.1 Two Explicit Approximations

The saddle-point density and distribution functions of the Hill estimator result from two explicit approximations:

- Under the second-order approximation (4.8), it follows that $H_{k,n} \overset{D}{\approx} \tilde{H}_{k,n}$.
- The density and distribution functions of $\tilde{H}_{k,n}$ are approximated by the saddle-point density and distribution functions respectively.

5.4.2 The Shape of the Saddle-point Density and Distribution Functions

The saddle-point density function is a continuous, smooth and non-negative function that decreases to 0 in the tails. Although the saddle-point density does not generally integrate to 1, its form implies that it can be renormalised by simply dividing by a constant. The value of the constant is determined by integrating the saddle-point density over its entire domain as is discussed in section 3.2.5. The renormalised saddle-point density function is a probability density function.

The saddle-point distribution function is a continuous, smooth and monotone-increasing function with the range $[0, 1]$, making it a cumulative distribution function. If the saddle-point $\hat{t}$ does not have a closed-form solution and has to be calculated numerically, then the resulting saddle-point distribution function can have a few discontinuities around the mean. These discontinuities can be eliminated by interpolating over them. Eliminating the saddle-point distribution function's discontinuities around the mean, which result from a numerically calculated saddle-point, does introduce a degree of complication. For this reason, the saddle-point density function is preferable to work with.

5.4.3 The Behaviour of the Saddle-point Approximation when Parameter Values are Estimated

In section 4.6 the behaviour of the saddle-point density and distribution functions of the Hill estimator are investigated when its parameters $\gamma, b_{n,k}$ and ρ are estimated. The results illustrate that both the saddle-point density and distribution functions are well-behaved in scenarios where the parameter estimates are unstable and inaccurate. In every case in which the parameters were estimated, both the saddle-point density and distribution functions were non-degenerate and well-behaved probability density and cumulative distribution functions respectively.

Furthermore, the Hill estimates and the saddle-point density and distribution function means and medians are all very similar when the parameters are estimated using maximum likelihood. The maximum likelihood estimated saddle-point density and distribution functions are the more stable and accurate than the regression estimated saddle-point density and distribution functions, even though the maximum likelihood parameter estimates are more unstable

and less accurate than the corresponding regression parameter estimates. This result suggests that there might be some underlying mechanism at play which makes the maximum likelihood saddle-point approximation accurate and stable while the corresponding maximum likelihood estimates are inaccurate and unstable.

5.5 Two Approaches to Deriving the Saddle-point Approximation

In the paper by Daniels (1954) in which he introduces the saddle-point approximation into statistics, he makes it clear that the saddle-point approximation can be derived using two methods: the saddle-point method and the indirect Edgeworth expansion. Barndorff-Nielsen & Cox (1979) derive the saddle-point approximation by using exponential tilting together with the Edgeworth expansion. Lugannani & Rice (1980) rely on a saddle-point method similar to that of Bleistein (1966) to derive their asymptotic expansion for the sample mean.

The two principal ways in which the saddle-point approximation is derived is by either using the saddle-point method or the indirect Edgeworth expansion. Each approach offers its own insights into the saddle-point approximation. The saddle-point method, however, has a wider range of applications than does the Edgeworth expansion.

5.5.1 Saddle-point Method

The saddle-point method, which is discussed in section 2.3, entails choosing the path of integration to traverse a saddle-point in such a way that the integral is focused in a small neighbourhood around the saddle-point. There are, however, various saddle-point methods, depending on the form of the function being integrated. The saddle-point method of de Bruijn (1958), when applied to the Fourier inversion formula of the density function of the sample mean, results in the uniform asymptotic expansion given by equation (3.5). The saddle-point method of Bleistein (1966), when applied to the Fourier inversion formula of the distribution function of the sample mean, results in the uniform asymptotic expansion given by equation (3.12).

The saddle-point method of Flajolet & Sedgewick (2009) can be applied in a more general setting than that of de Bruijn (1958) due to the weaker conditions placed on the integral (see section 2.3.5). The saddle-point approximation that results from applying the saddle-point method of Flajolet & Sedgewick (2009) to a Fourier inversion formula of a density function can be applied to approximate any density function whose corresponding cumulant generating function satisfies the imposed conditions. This saddle-point approximation can therefore be used in a more general setting and yields an asymptotic result.

5.5.2 Indirect Edgeworth Expansion

The Edgeworth expansion can be used indirectly to yield a uniform asymptotic expansion of the distribution of the sample mean. Daniels (1954) is the first to prove how the Edgeworth expansion can be used indirectly to yield a uniform asymptotic estimate. He manipulates the Fourier inversion formula of the density function of the sample mean so that it is the product of two terms, the latter being the density function of a sample mean at the point 0, where both terms depend on a parameter t. By approximating the latter with its Edgeworth expansion and choosing t as the saddle-point $\hat{t}$, the saddle-point approximation is derived.

Barndorff-Nielsen & Cox (1979) also use the Edgeworth expansion indirectly to derive the saddle-point. They, however, use a method called exponential tilting in order to apply the Edgeworth expansion indirectly. Exponential tilting simply entails embedding the density function of interest $f(x)$ into a conjugate family of distributions $f(x;t)$ as follows:

$$f(x;t) = e^{xt-K(t)}\, f(x)$$

where $K(t)$ is the cumulant generating function corresponding to the density function $f(x)$. Note that the density function $f(x;t)$ is in the exponential family and that $f(x;0) = f(x)$. A complete derivation of the saddle-point approximation using exponential tilting together with the Edgeworth expansion is given by Barndorff-Nielsen & Cox (1979:280-281).

5.6 Areas of Further Research

There are three areas of further research to highlight:

The first is the need for rigorous theory to be developed to support the high accuracy and good behaviour of the saddle-point approximation in a non-asymptotic setting. In this thesis the two conditions (3.2) and (3.7) are proposed and are expected to yield the saddle-point density and distribution functions of Butler (2007). However, these adaptations have not been proved and their main aim is to highlight the area in which the saddle-point method theory differs from the saddle-point density and distribution formulas of Butler (2007). Furthermore, the implications of these adaptations are not clear. They have been proposed simply in an attempt to link the saddle-point method theory to the formulas of Butler (2007).

The second is the estimation of the saddle-point density and distribution functions. The saddle-point density and distribution functions are estimated in section 4.6, and this yields good results, especially considering that a sample of size 10 is chosen. It would be interesting to investigate how the estimated saddle-point density and distribution functions behave for other random variables, and which estimation methods yield the best results.

The third is the possibility of determining the distribution of a random variable by explicitly defining its saddle-point. The saddle-point density and distribution functions are explicit functionals of the saddle-point. Therefore, by explicitly defining the saddle-point appropriately, the corresponding cumulant generating function can be found, and using these two in conjunction would yield the saddle-point density and distribution functions.

Chapter 6

References

Akkouchi, M. 2008. On the Convolution of Exponential Distributions. *Journal of the Chungcheong Mathematical Society*, 21(4):501-510.

Balkema, A. & de Haan, L. 1974. Residual Life Time at Great Age. *Annals of Probability*, 2(5):792-804.

Barndorff-Nielsen, O. E. & Cox, D. R. 1979. Edgeworth and Saddlepoint Approximations with Statistical Applications. *Journal of the Royal Statistical Society: Series B*, 41:279-312.

Barndorff-Nielsen, O. E. 1983. On a Formula for the Distribution of the Maximum Likelihood Estimator. *Biometrika*, 70:343-365.

Bleistein, N. 1966. Uniform Asymptotic Expansions of Integrals with Stationary Point Near Algebraic Singularity. *Communications on Pure and Applied Mathematics*, 19(4):353-370.

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. 2004. *Statistics of Extremes: Theory and Applications*. Chichester: John Wiley & Sons Ltd.

Berry, M. V. 2016. Riemann's Saddle-point Method and the Riemann-Siegel Formula, in Ji, L., Oort, F. & Yau, S. T. (eds.) *The Legacy of Bernhard Riemann After One Hundred and Fifty Years*. Boston: International Press of Boston, Inc. 69-78.

Butler, R. 2007. *Saddlepoint Approximations with Applications*. Cambridge: Cambridge University Press.

Casella, G. & Goutis, C. 1999. Explaining the Saddlepoint Approximation. *The American Statistician*, 53(3):216-224.

Cheng, S. & Pan, J. 1998. Asymptotic Expansions of Estimators for the Tail Index with Applications. *Scandinavian Journal of Statistics*, 25:717-728.

Chen, Q. & Giles, D. E. 2008. General Saddlepoint Approximations: Applications to the Anderson-Darling Test Statistic. *Communications in Statistics - Simulation and Computation*, 37(4):789-804.

Conway, J. B. 1973. *Functions of One Complex Variable*. New York: Springer-Verlag.

Cuntz, A., Haeuslter, E. & Segers, J. J. J. 2003. Edgeworth Expansions for the Distribution Function of the Hill Estimator. *Center of Economic Research: Discussion Paper*, 2003(8). Tilburg: Econometrics.

Daniels, H. E. 1954. Saddlepoint Approximation in Statistics. *Annals of Mathematical Statistics*, 25:631-650.

Daniels, H. E. 1982. The Saddlepoint Approximation for a General Birth Process. *Journal of Applied Probability*, 19(1):20-28.

Daniels, H. E. 1983. Saddlepoint Approximations for Estimating Equations. *Biometrika*, 70(1):89-96.

Daniels, H. E. 1987. Tail Probability Approximations. *International Statistical Review*, 55(1):37-43.

Daniels, H. E. 1988. Comment (on article by N. Reid). *Statistical Science*, 3(2):229.

Davison, A. C. & Hinkley, D. V. 1988. Saddlepoint Approximations in Resampling Methods. *Biometrika*, 75(3):417-431.

Debye, P. 1909. Naherungsformeln fur die Zylinderfunktionen fur groe Werte des Arguments und unbeschrankt veranderliche Werte des Index. *Mathematische Annalen*, 67:535-558.

de Bruijn, N. G. 1958. *Asymptotic Methods in Analysis*. Amsterdam: North-Holland.

Easton, G. S. & Ronchetti, E. 1986. General Saddlepoint Approximations with Applications to L-statistics. *Journal of the American Statistical Association*, 81(394):420-430.

Field, C. A. & Ronchetti, E. 1990. *Small Sample Asymptotics*. Hayward: Institute of Mathematical Statistics.

Field, C. A. & Hampel, F. R. 1982. Small-sample Asymptotic Distributions of M-estimators of Location. *Biometrika*, 69(1):29-46.

Fisher, R. A. & Tippett, L. H. C. 1928. On the Estimation of the Frequency Distributions of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24:180-190.

Flajolet, P. & Sedgewick, R. 2009. *Analytic Combinatorics*. Cambridge: Cambridge University Press.

Gatto, R. 2010. A Saddlepoint Approximation to the Distribution of Inhomogeneous Discounted Compound Poisson Processes. *Methodology and Computing in Applied Probability*, 12(3):533-551.

Gatto, R. 2008. A Saddlepoint Approximation to the Probability of Ruin in the Compound Poisson Process with Diffusion. *Statistics & Probability Letters*, 78(13):1948-1954.

Giles, D. E. A. 2001. A Saddlepoint Approximation to the Distribution of the Anderson-Darling Test Statistic. *Communications in Statistics - Simulation and Computation*, 30(4):899-905.

Gnedenko, B. 1943. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44(3):423-453.

Hougaard, P. 1988. Comment (on article by N. Reid). *Statistical Science*, 3(2):230-231.

Lopez, J. L., Pagola, P. & Sinusia, E. P. 2009. A Systematization of the Saddle Point Method. Application to the Airy and Hankel Functions. *Journal of Mathematical Analysis and Applications*, 354(1):347-359.

Lugannani, R. & Rice, S. 1980. Saddle Point Approximation for the Distribution of the Sum of Independent Random Variables. *Advances in Applied Probability*, 12(2):475-490.

Lukacs, E. 1970. *Characteristic Functions*. London: Charles Griffin & Company Ltd.

Murakami, H. 2009. Saddlepoint Approximations to the Limiting Distribution of the Modified Anderson-Darling Test Statistic. *Communications in Statistics - Simulation and Computation*, 38(10):2214-2219.

Nekrasov, P. A. 1886. The Lagrange Series and Approximate Expressions for Functions of Very Large Numbers. *Matematicheski Sbornik*, 12:481-579.

Petrova, S. S. & Solov'ev, A. D. 1997. The Origin of the Method of Steepest Descent. *Historia Mathematica*, 24:361-375.

Pickands, J. 1975. Statistical Inference using Extreme Order Statistics. *Annals of Statistics*, 3(1):119-131.

Reid, N. 1988. Saddlepoint Methods and Statistical Inference. *Statistical Science*, 3(2):213-238.

Rogers, L. C. G. & Zane, O. 1999. Saddlepoint Approximations to Option Prices. *The Annals of Applied Probability*, 9(2):493-503.

Renshaw, E. 2000. Applying the Saddlepoint Approximation to Bivariate Stochastic Processes. *Mathematical Biosciences*, 168(1):57-75.

Spady, R. H. 1991. Saddlepoint Approximations for Regression Models. *Biometrika*, 78(4):879-889.

Siegel, C. L. 1932, Uber Riemanns Nachlass zur analytisches Zahlentheorie. *Quellen und Studien zur Geschichte der Mathematik, Astronomie, und Physik*, 2:45-80.

Wood, A. T. A., Booth, J. G. & Butler, R. W. 1993. Saddlepoint Approximations to the CDF of Some Statistics with Nonnormal Limit Distributions. *Journal of the American Statistical Association*, 88(422):680-686.

Chapter 7

Appendix: R Code

7.1 R Code for Chapter 3

```
# saddlepoint density function of non-central chi-squared
p = 7
l = 5

x = seq(.01,50,by=.01)
plot(x=x, y=dchisq(x,df=p, ncp=l), type="l", ylab="density",
     main="Non-central Chi-squared Density with ncp=5 and df=7")

r=(p+sqrt(p^2+4*l*x))/(2*x)
saddle = r^(1-p/2)*exp((1-r)*(l/r-x)/2)/2/sqrt(pi)/sqrt(2*l/r+p)
lines(x=x, y=saddle, lty=2, col=2)

# normalising constant
nc=1/sum(saddle*.01)
lines(x=x, y=nc*saddle, lty=2, col=3)


# saddlepoint distribution function of non-central chi-squared
p = 8
l = 5
x = seq(.01,40,by=.01)

# sp is the old saddlepoint
sp = (2*x-p-sqrt(p^2+4*l*x))/(4*x)

# Lugannani Rice amounts
w = sign(sp)*sqrt(2)*sqrt(sp*x-l*sp/(1-2*sp)+p/2*log(1-2*sp))
u = sp*sqrt(4*l/(1-2*sp)^3+2*p/(1-2*sp)^2)

sp_lr = pnorm(w)-dnorm(w)*(1/u-1/w)

plot(x=x, y=pchisq(x,df=p, ncp=l), type="l", ylab="density",
     main="Non-central Chi-squared Distribution Function with
               ncp=5 and df=7")
lines(x=x, y=sp_lr, lty=2, col=4)
```

7.2 R Code for Chapter 4

7.2.1 The sp_density function

```
sp_density = function(n=100, k=10, m=20, dist=rburr, lim=1,
                      true_p=NULL, seed=2, plothill=F, ...)
{
  domain=seq(.01,lim,length.out=1000)

  if(m>0)
  {
    # create m samples, each of size n,
    # and determine the corresponding m samples of z, each of size k
    set.seed(seed)
    x = matrix(dist(n*m, ...), nrow=n, ncol=m)
    z = apply(x,2,function(arg){
      (1:k)*(log(sort(arg,decreasing=T))[1:k]
      -log(sort(arg,decreasing=T))[2:(k+1)])})



    # MAXIMUM LIKELIHOOD ESTIMATES

    # determine the mle's of gamma, rho and bnk for each of the m samples
    mle = apply(z,2,function(arg){
      optim(par=c(0,0,0), fn=function(par,dat=arg){
        l = exp(par[1])+((1:k)/k)^(exp(par[2]))*exp(par[3])
        return(sum(log(l))+sum(dat/l))
      })$par
    })

    # determine the mle's of the lambdas (of which there are k)
    # for each of the m samples
    mle_l = apply(mle,2,function(arg){
      exp(arg[1])+((1:k)/k)^(exp(arg[2]))*exp(arg[3])})

    # determine the mle's of the saddle-points for each of the m samples
    mle_sp = apply(mle_l,2,function(arg){
      if(max(arg)-min(arg)<10^(-10))
        k/min(arg)-k/domain

      else
        sapply(domain, function(arg1){
          uniroot(f=function(t){
            sapply(t,function(arg2){sum(arg/(k-arg*arg2))})-arg1},
            interval=c(k/max(arg)-k/arg1,
                       min(k/max(arg), k/min(arg)-k/arg1)))$root
        })
    })
```

```
# determine the numerator and denominator of the
# mle saddle-point density function for each of the m samples

mle_num = sapply(1:m, function(i){
  exp(sapply(mle_sp[,i],function(arg){
    -sum(log(1-mle_l[,i]*arg/k))})-mle_sp[,i]*domain)})

mle_denum = sapply(1:m, function(i){
  sqrt(2*pi)*sqrt(sapply(mle_sp[,i],function(arg){
    sum(mle_l[,i]^2/(k-mle_l[,i]*arg)^2)}))})


# the Edgeworth expansion of the mle estimates
mle_edge = apply(mle,2,function(arg){
  temp = (domain-exp(arg[1]))*sqrt(k)/exp(arg[1])
  return(sqrt(k)/exp(arg[1])
         *dnorm(temp)*(1-temp/sqrt(k)+temp*sqrt(k)*exp(arg[3])
                       /exp(arg[1])/(1+exp(arg[2]))+temp^3/3/sqrt(k)))
})



# REGRESSION ESTIMATES

reg = apply(z,2,function(arg){
  temp = optimise(f=function(arg1){cor(arg,((1:k)/k)^exp(arg1))},
                  interval=c(-100,2), maximum=T)
  return(ifelse(temp$objective>0,temp$maximum,0))})

reg_l = sapply(1:m,function(i){
  temp = ((1:k)/k)^exp(reg[i])
  return(lm.fit(x=matrix(temp,ncol=1), y=z[,i])$fitted.values)
})

# determine the regression estimates of the saddle-points
# for each of the m samples

reg_sp = apply(reg_l,2,function(arg){
  if(max(arg)-min(arg)<10^(-10))
    k/min(arg)-k/domain

  else
    sapply(domain, function(arg1){
      uniroot(f=function(t){
        sapply(t,function(arg2){sum(arg/(k-arg*arg2))})-arg1},
        interval=c(k/max(arg)-k/arg1,min(k/max(arg),
                                         k/min(arg)-k/arg1)))$root
    })
})
```

```
  reg_num = sapply(1:m, function(i){
    exp(sapply(reg_sp[,i],function(arg){
      -sum(log(1-reg_l[,i]*arg/k))})-reg_sp[,i]*domain)})

  reg_denum = sapply(1:m, function(i){
    sqrt(2*pi)*sqrt(sapply(reg_sp[,i],function(arg){
      sum(reg_l[,i]^2/(k-reg_l[,i]*arg)^2)}))})


  # the Edgeworth expansion of the regression estimates
  reg_edge = sapply(1:m,function(i){
    covar = ((1:k)/k)^exp(reg[i])
    para = lm(z[,i]~covar)$coef
    temp = (domain-para[1])*sqrt(k)/para[1]
    return(sqrt(k)/para[1]
           *dnorm(temp)*(1-temp/sqrt(k)+temp*sqrt(k)*para[2]
                         /para[1]/(1+exp(reg[i]))+temp^3/3/sqrt(k)))
  })
}



# TRUE SADDLE POINT AND EDGEWORTH DENSITIES
if(!is.null(true_p))
{
  true_l = true_p[1]+((1:k)/k)^(-true_p[2])*true_p[3]
  true_sp = sapply(domain, function(arg1){
    uniroot(f=function(t){sapply(t,function(arg2){
      sum(true_l/(k-true_l*arg2))})-arg1},
      interval=c(k/max(true_l)-k/arg1,
                 min(k/max(true_l), k/min(true_l)-k/arg1)))$root})

  true_num = exp(sapply(true_sp,function(arg){
    -sum(log(1-true_l*arg/k))})-true_sp*domain)

  true_denum = sqrt(2*pi)*sqrt(sapply(true_sp,function(arg){
    sum(true_l^2/(k-true_l*arg)^2)}))

  temp = (domain-true_p[1])*sqrt(k)/true_p[1]
  temp = (domain-true_p[1])*sqrt(k)/true_p[1]
  true_edge = sqrt(k)/true_p[1]
              *dnorm(temp)*(1-temp/sqrt(k)+temp*sqrt(k)*true_p[3]
                            /true_p[1]/(1-true_p[2])+temp^3/3/sqrt(k))
}
```

```
# NOW TO PLOT THEM ALL!
if(m>0)
{
  par(mfrow=c(2,1))

  plot(x=c(0,lim), y=c(-.1,max(mle_num/mle_denum,reg_num/reg_denum)),
       type="n", ylab="Density", xlab="x",
       main=paste("Maximum Likelihood Estimated Density Functions (n=",
                  n,", k=",k,")",sep=""))

  if(plothill==T)
    abline(v=hill, lty=2)

  for(i in 1:m){
    lines(x=domain, col="lightgreen", lwd=1.1,
          y=mle_num[,i]/mle_denum[,i]/sum(mle_num[,i]/mle_denum[,i])/.01)
    lines(x=domain, y=mle_edge[,i], col="tomato", lwd=1)
  }

  if(!is.null(true_p))
  {
    lines(lwd=2, col=3, x=domain,
          y=true_num/true_denum/sum(true_num/true_denum)/.01)
    lines(x=domain, y=true_edge, lwd=2, col=2)
  }


  plot(x=c(0,lim), y=c(-.1,max(reg_num/reg_denum,mle_num/mle_denum)),
       type="n", ylab="Density", xlab="x",
       main=paste("Regression Estimated Density Functions (n=",
                  n,", k=",k,")",sep=""))

  if(plothill==T)
    abline(v=hill, lty=2)

  for(i in 1:m){
    lines(col="lightgreen", lwd=1.1, x=domain,
          y=reg_num[,i]/reg_denum[,i]/sum(reg_num[,i]/reg_denum[,i])/.01)
    lines(x=domain, y=reg_edge[,i], col="tomato", lwd=1)
  }

  if(!is.null(true_p))
  {
    lines(lwd=2, col=3, x=domain,
          y=true_num/true_denum/sum(true_num/true_denum)/.01)
    lines(x=domain, y=true_edge, lwd=2, col=2)
  }
  par(mfrow=c(1,1))
}
```

```
  if(m==0)
  {
    set.seed(seed)
    hill = apply(matrix(log(dist(100000*n, ...)), ncol=n), 1, function(arg){
      mean(sort(arg,decreasing=T)[1:k])-sort(arg,decreasing=T)[k+1]
    })

    hist(hill, freq=F, breaks=70, xlab="Hill Estimate",
         main="Histogram of the Hill Estimator")
    lines(x=domain, y=true_num/true_denum, lwd=2, col=3)
    lines(x=domain, y=true_edge, lwd=2, col=2)
  }

}

library(actuar)

n = 100
k = 10

# Burr
a = 1/2
b = 2/3
sp_density(dist=rburr, shape1=a, shape2=b, lim=3/a/b, m=0,
           true_p=c(1/(a*b), -1/a, 1/(a*b)/((n/k)^(1/a)-1)))

sp_density(dist=rburr, shape1=a, shape2=b, lim=3/a/b,
           true_p=c(1/(a*b), -1/a, 1/(a*b)/((n/k)^(1/a)-1)))

# Frechet
a=1/3
sp_density(dist=function(arg,shape1){(-log(runif(arg)))^(-1/shape1)},
           shape1=a, m=0, lim=3/a, true_p=c(1/a, -1, k/(2*a*n)))

sp_density(dist=function(arg,shape1){(-log(runif(arg)))^(-1/shape1)},
           shape1=a, lim=3/a, true_p=c(1/a, -1, k/(2*a*n)))

# Inverse Gamma
sp_density(dist=rinvgamma, shape=1/3, rate=1, lim=9)

# Log Gamma
sp_density(dist=rlgamma, shapelog=1, ratelog=1/3, lim=9)
```

7.2.2 The sp_dist function

```
sp_dist = function(n=100, k=10, m=20, dist=rburr, lim=1, r=.1,
                   true_p=NULL, seed=2, plothill=F, ...)
{
  domain=seq(.01,lim,length.out=1000)

  if(m>0)
  {
    # create m samples, each of size n,
    # and determine the corresponding m samples of z, each of size k
    set.seed(seed)
    x = matrix(dist(n*m, ...), nrow=n, ncol=m)
    z = apply(x,2,function(arg){
      (1:k)*(log(sort(arg,decreasing=T))[1:k]
      -log(sort(arg,decreasing=T))[2:(k+1)])})



    # MAXIMUM LIKELIHOOD ESTIMATES

    # determine the mle's of gamma, rho and bnk for each of the m samples
    mle = apply(z,2,function(arg){
      optim(par=c(0,0,0), fn=function(par,dat=arg){
        l = exp(par[1])+((1:k)/k)^(exp(par[2]))*exp(par[3])
        return(sum(log(l))+sum(dat/l))
      })$par
    })

    # determine the mle's of the lambdas (of which there are k)
    # for each of the m samples
    mle_l = apply(mle,2,function(arg){
      exp(arg[1])+((1:k)/k)^(exp(arg[2]))*exp(arg[3])})

    # determine the mle's of the saddle-points for each of the m samples
    mle_sp = apply(mle_l,2,function(arg){
      if(max(arg)-min(arg)<10^(-10))
        k/min(arg)-k/domain

      else
        sapply(domain, function(arg1){
          uniroot(f=function(t){
            sapply(t,function(arg2){sum(arg/(k-arg*arg2))})-arg1},
            interval=c(k/max(arg)-k/arg1,
                       min(k/max(arg), k/min(arg)-k/arg1)))$root
        })
    })
```

```
# determine the Lugannani-Rice amounts for the mles
mle_w = sapply(1:m, function(i){
  sign(mle_sp[,i])
  *sqrt(2)*sqrt(mle_sp[,i]*domain+sapply(mle_sp[,i],function(t){
    sum(log(1-mle_l[,i]/k*t))}))})
mle_u = sapply(1:m, function(i){
  mle_sp[,i]*sqrt(sapply(mle_sp[,i],function(t){
    sum(mle_l[,i]^2/(k-mle_l[,i]*t)^2)}))})

# the Edgeworth expansion of the mle estimates
mle_edge = apply(mle,2,function(arg){
  temp = (domain-exp(arg[1]))*sqrt(k)/exp(arg[1])
  return(pnorm(temp)+dnorm(temp)
         *((1-temp^2)/3/sqrt(k)-sqrt(k)*exp(arg[3])
           /exp(arg[1])/(1+exp(arg[2]))))
})



# REGRESSION ESTIMATES

reg = apply(z,2,function(arg){
  temp = optimise(f=function(arg1){cor(arg,((1:k)/k)^exp(arg1))},
                  interval=c(-100,2), maximum=T)
  return(ifelse(temp$objective>0,temp$maximum,0))})

reg_l = sapply(1:m,function(i){
  temp = ((1:k)/k)^exp(reg[i])
  return(lm.fit(x=matrix(temp,ncol=1), y=z[,i])$fitted.values)
})

# determine the regression estimates of the saddle-points
# for each of the m samples

reg_sp = apply(reg_l,2,function(arg){
  if(max(arg)-min(arg)<10^(-10))
    k/min(arg)-k/domain

  else
    sapply(domain, function(arg1){
      uniroot(f=function(t){
        sapply(t,function(arg2){sum(arg/(k-arg*arg2))})-arg1},
        interval=c(k/max(arg)-k/arg1,min(k/max(arg),
                                         k/min(arg)-k/arg1)))$root
    })
})
```

```
  # determine the Lugannani-Rice amounts for the regression estimates
  reg_w = sapply(1:m, function(i){
    sign(reg_sp[,i])*sqrt(2)
    *sqrt(reg_sp[,i]*domain+sapply(reg_sp[,i],function(t){
      sum(log(1-reg_l[,i]/k*t))}))})

  reg_u = sapply(1:m, function(i){
    reg_sp[,i]*sqrt(sapply(reg_sp[,i],function(t){
      sum(reg_l[,i]^2/(k-reg_l[,i]*t)^2)}))})

  # the Edgeworth expansion of the regression estimates
  reg_edge = sapply(1:m,function(i){
    covar = ((1:k)/k)^exp(reg[i])
    para = lm(z[,i]~covar)$coef
    temp = (domain-para[1])*sqrt(k)/para[1]
    return(pnorm(temp)+dnorm(temp)
           *((1-temp^2)/3/sqrt(k)-sqrt(k)*para[2]
             /para[1]/(1+exp(reg[i]))))
  })

}



# TRUE SADDLE POINT AND EDGEWORTH DISTRIBUTION FUNCTIONS
if(!is.null(true_p))
{
  true_l = true_p[1]+((1:k)/k)^(-true_p[2])*true_p[3]
  true_sp = sapply(domain, function(arg1){
    uniroot(f=function(t){sapply(t,function(arg2){
      sum(true_l/(k-true_l*arg2))})-arg1},
      interval=c(k/max(true_l)-k/arg1,
                 min(k/max(true_l), k/min(true_l)-k/arg1)))$root})

  # determine the Lugannani-Rice amounts
  true_w = sign(true_sp)*sqrt(2)
           *sqrt(true_sp*domain+sapply(true_sp,function(t){
             sum(log(1-true_l/k*t))}))
  true_u = true_sp*sqrt(sapply(true_sp,function(t){
      sum(true_l^2/(k-true_l*t)^2)}))

  temp = (domain-true_p[1])*sqrt(k)/true_p[1]
  true_edge = pnorm(temp)
              +dnorm(temp)*((1-temp^2)/3/sqrt(k)-sqrt(k)
                            *exp(true_p[3])/exp(true_p[1])
                            /(1-true_p[2]))
}
```

```
# NOW TO PLOT THEM ALL!
if(m>0)
{
  par(mfrow=c(2,1))

  plot(x=c(0,lim), y=c(-.1,1.1), type="n",
       main=paste("MLE Distribution Functions (n=",
                  n,", k=",k,")",sep=""),
       ylab="Cumulative Distribution", xlab="x")

  for(i in 1:m){
    ind = abs(domain-mean(mle_l[,i]))<r
    lines(col="lightblue", lwd=1.1, x=domain[!ind],
          y=(pnorm(mle_w[,i])-dnorm(mle_w[,i])*(1/mle_u[,i]
                                                -1/mle_w[,i]))[!ind])
    lines(x=domain, y=mle_edge[,i], col="tomato", lwd=1)
  }

  if(!is.null(true_p))
  {
    ind = abs(domain-mean(true_l))<r
    lines(lwd=2, col=4, x=domain[!ind],
          y=(pnorm(true_w)-dnorm(true_w)*(1/true_u-1/true_w))[!ind])
    lines(x=domain, y=true_edge, lwd=2, col=2)
  }

  plot(x=c(0,lim), y=c(-.1,1.1), type="n",
       main=paste("Regression Estimated Distribution Functions (n=",
                  n,", k=",k,")",sep=""),
       ylab="Cumulative Distribution", xlab="x")

  for(i in 1:m){
    ind = abs(domain-mean(reg_l[,i]))<r
    lines(col="lightblue", lwd=1.1, x=domain[!ind],
          y=(pnorm(reg_w[,i])-dnorm(reg_w[,i])*(1/reg_u[,i]
                                                -1/reg_w[,i]))[!ind])
    lines(x=domain, y=reg_edge[,i], col="tomato", lwd=1)
  }

  if(!is.null(true_p))
  {
    ind = abs(domain-mean(true_l))<r
    lines(lwd=2, col=4, x=domain[!ind],
          y=(pnorm(true_w)-dnorm(true_w)*(1/true_u-1/true_w))[!ind])
    lines(x=domain, y=true_edge, lwd=2, col=2)
  }
  par(mfrow=c(1,1))
}
```

```
  if(m==0)
  {
    set.seed(seed)
    hill = apply(matrix(log(dist(100000*n, ...)), ncol=n), 1, function(arg){
      mean(sort(arg,decreasing=T)[1:k])-sort(arg,decreasing=T)[k+1]
    })

    plot(ecdf(hill),
         main="Empirical Distribution Function of the Hill Estimator",
         ylab="Cumulative Probability")
    ind = abs(domain-mean(true_l))<r
    lines(lwd=2, col=4, x=domain[!ind],
          y=(pnorm(true_w)-dnorm(true_w)*(1/true_u-1/true_w))[!ind])
    lines(x=domain, y=true_edge, lwd=2, col=2)
  }

}

library(actuar)

n = 100
k = 10

# Burr
a = 1/2
b = 2/3
sp_dist(dist=rburr, shape1=a, shape2=b, lim=3/a/b, m=0,
        true_p=c(1/(a*b), -1/a, 1/(a*b)/((n/k)^(1/a)-1)))

sp_dist(dist=rburr, shape1=a, shape2=b, lim=3/a/b,
        true_p=c(1/(a*b), -1/a, 1/(a*b)/((n/k)^(1/a)-1)))

# Frechet
a=1/3
sp_dist(dist=function(arg,shape1){(-log(runif(arg)))^(-1/shape1)},
        shape1=a, lim=3/a, m=0,
        true_p=c(1/a, -1, k/(2*a*n)))

sp_dist(dist=function(arg,shape1){(-log(runif(arg)))^(-1/shape1)},
        shape1=a, lim=3/a,
        true_p=c(1/a, -1, k/(2*a*n)))

# Inverse Gamma
sp_dist(dist=rinvgamma, shape=1/3, rate=1, lim=9)

# Log Gamma
sp_dist(dist=rlgamma, shapelog=1, ratelog=1/3, lim=9)
```